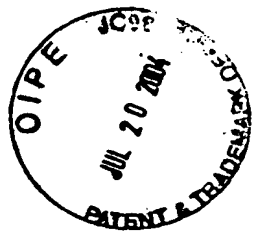




AAAAATAAATCAT ATG AAA AAT ATT AAA AAA AAT CAA GTA ATC AAT CTC GGT CCT AAT TCT  
M K N I K K N Q V M N L G P N S  
AAA TTA TTA AAA GAA TAT AAA TCA CAA TTA ATT GAA TTA AAT ATT GAA CAA TTT GAA GCA  
K L L K E Y K S Q L I I E L N I E Q F E A  
GGT ATT GGT TTA ATT TTA GGA GAT GCT TAT ATT CGT AGT CGT GAT GAA GGT AAA ACT TAT  
G I G L I L G D A Y I R S R D E G K T Y  
TGT ATG CAA TTT CAC TCC AAA AAT AAG GCA TAC ATG GAT CAT GTA TGT TTA TTA TAT GAT  
C M Q F E W K N K A Y M D H V C L L Y D  
CAA TGG GTA TTA TCA CCT CCT CAT AAA AAA GAA AGA GTT AAT CAT TTA GGT AAT TTA GTA  
Q W V L S C P P H K K E R V N H L G N L V  
ATT ACC TGG GGA GCT CAA ACT TTT AAA CAT CAA GCT TTT AAT AAA TTA GCT AAC TTA TTT  
I T W G A Q T F K H Q A F N K L A N L F  
ATT GTA AAT AAT AAA CTT ATT CCT AAT AAT TTA GTT GAA AAT TAT TTA ACA CCT ATG  
I Y N N K K L I P N N L V E N Y L T P M  
AGT CTG GCA TAT TGG TTT ATG GAT GAT GGA GGT AAA TGG GAT TAT AAT AAA AAT TCT CTT  
S L A Y W F M D D G G K W D Y N K N S L  
AAT AAA AGT ATT GTA TTA AAT ACA CAA AGT TTT ACT TTT GAA GAA GTA GAA TAT TTA CTT  
N K S I V L N T Q S F T F E E V C Y L V  
AAA GGT TTA AGA AAT AAA TTT CAA TTA AAT TGT TAT GTT AAA ATT AAT AAA AAT AAA CCA  
K G L R N K F Q L N C Y V K I N K N K P  
ATT ATT TAT ATT GAT TCT AGT AGT TAT CTG ATT TTT TAT AAT TTA ATT AAA CCT TAT TTA  
I I Y I D S M S Y L I F Y N I T K P Y L  
ATT CCT CAA ATG ATG TAT AAA CTG CCT AAT ACT ATT TCA TCC GAA ACT TTT TTA AAA TAA  
I P Q M Y K L P N T I S S E T F L K

FIG. 1



Bam HI

CCGGATCCATG CAT ATG AAA AAC ATC AAA AAC CAG GTA ATG AAC CTG GGT CCG AAC TCT  
 AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC ATC GAA CAG TTC GAA GCA  
 GGT ACT GGT CTG ATC CTG GGT GAT TAC ATC CGT TCT CGT GAT GAA GGT AAA ACC TAC  
 TGT ATG CAG TTC GAG TGG AAA AAC GCA TAC ATG GAC CAC GTA TGT CTG CTG TAC GAT  
 CAG TGG GTA CTG TCC CCG CCG CAC AAA GAA GAA GGT GGT AAC CTG GTA  
 ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC AAA CTG GCT AAC CTG TTC  
 ATC GTT AAC AAC AAA ACC ATC CCG AAC AAC CTG GTT GAA AAC TAC CTG ACC CCG ATG  
 TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GAT TAC AAC AAA AAC TCT ACC  
 AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA GAA GTA TAC CTG GTT  
 AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC AAA ATC AAC AAA AAC CCG  
 ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC AAC CTG ATC AAA CCG TAC CTG  
 ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC GAA ACT TTC CTG AAA TAA  
 TAAGTCGACTGCAGGATCCGGTAAGTAAGTAA

Sall PstI BamHI

1 and 2: THESE AMINO ACIDS ARE ABSOLUTELY NECESSARY TO PRODUCE CATALYTIC  
 ACTIVITY. OTHER SUBSTITUTIONS ARE POSSIBLE, SUCH AS DELETIONS  
 OF THE 10 FIRST AMINO ACIDS.

FIG. 2

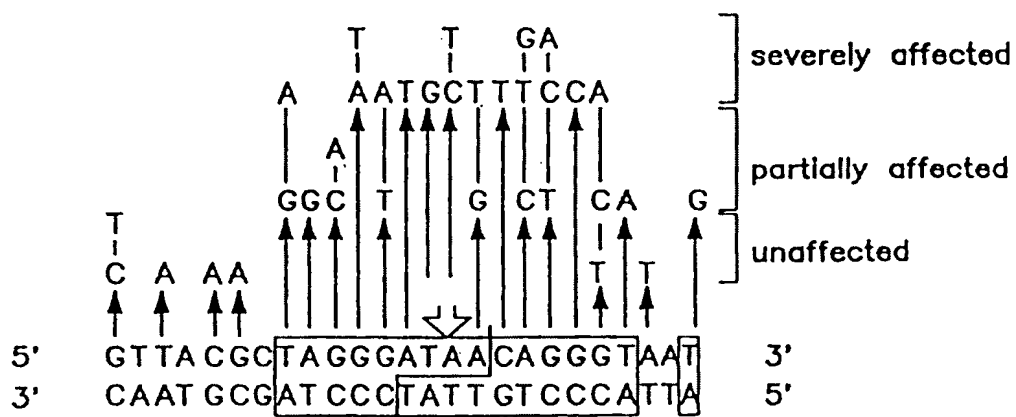


FIG. 3



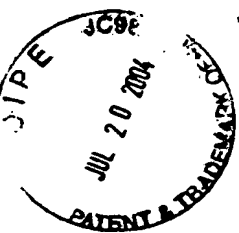
1667 GCGGACAGGTATCCGGTAAGCGGCGAGGGTCGGAACAGGAGCGCAGGGAGCTTCCAGGGGGAACGCCTGGTATCT 1746  
1747 TTATAGTCCGTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTG ATG CTC GTC AGG GGG GCG GAG 1818  
1819 CCT ATG GAA AAA CGC CAG CAA CGC GGC CTT TTT ACG GTT CCT GGC CTT TTG CTG GCC TTT 1878  
1879 TGC TCA CAT GTT CTT TCC TGC TGT ATC CCC TGA TTCGTGGATAACCGTATTACCGCCCTTTGAGTGAGC 1947  
1948 TGATACCGCTCGCGCAGCCGAACGACGAGCGCAGGAGTCAGTGAGCGAGGAGCGGCCCAATACGCAAC 2027  
2028 CGCCTCTCCCCGCGGTTGGCCGATTCAATTA ATG CAG CTG GCA CGA CAG GTT TCC CGA CTG GAA AGC 2094  
2095 GGG CAG TGA GCGCAACGCAATTA ATG TGA GTTAGCTCACTCATTAGGCACCCCGGCTTTACACITTT ATG 2164  
2165 CTT CCG GCT CGT ATG TTG TGT GGA ATT GTG AGC GGA TAA CAATTTTCACACAGGAACAGCT ATG 2228  
2229 ACC ATG ATT ACG AAT TCT CAT GTT TGA CAGCTTATCATCGATAAGCTTTA ATG CGG TAG TTTATCAC 2295  
2296 AGTTAAATTGCTAACGCAGTCAGGCACCGTGT ATG AAA TCT AAC AAT GCG CTC ATC GTC ATC CTC GGC 2363  
2364 ACC GTC ACC CTG GAT GCT GTA GGC ATA GGC TTG GTT ATG CCG GTA CTG CCG GGC CTC TTG 2423  
2424 CGG GAT ATC CGC CTG ATG CGT GAA CGT GAC GGA CGT AAC CAC CGC GAC ATG TGT GTG CTG 2483  
2484 TTC CGC TGG GCA TGC CAG GAC AAC TTC TGG TCC GGT AAC GTG CTG AGC CCG GCC AAG CTT 2543

FIG. 4A



2544	ACT	CCC	CAT	CCC	CCT	GTT	GAC	AAT	IAA	TCATCGGCTCGTATA	ATG	TGT	GGA	ATT	GTG	AGC	GGA	2606
73	T	P	H	P	P	V	D	N				C	G	I	V	S	G	7
2607	TAA	CAATTT	CACACAGGAAACAGGATCC	BamHI														2670
8				M	H	M	K	N	AAC	ATC	AAA	AAA	AAA	AAA	AAA	AAA	AAA	12
2671	AAC	CTG	GGT	CCG	AAC	TCT	AAA	CTG	CTG	AAA	GAA	TAC	AAA	TCC	CAG	CTG	ATC	2730
13	N	L	G	P	P	N	S	K	L	K	E	Y	K	S	Q	L	I	32
2731	ATC	GAA	CAG	TTC	GAA	GCA	GGT	ATC	GGT	CTG	ATC	CTG	GGT	GAT	GCT	TAC	ATC	2790
33	I	E	Q	F	E	A	G	I	G	L	I	L	G	D	A	Y	I	52
2791	GAT	GAA	GGT	AAA	ACC	TAC	TGT	ATG	CAG	TTC	GAG	TGG	AAA	AAC	AAA	GCA	TAC	2850
53	D	E	G	K	T	Y	C	M	Q	F	E	W	K	N	K	A	Y	72
2851	GTA	TGT	CTG	CTG	TAC	GAT	CAG	TGG	GTA	CTG	TCC	CCG	CAC	AAA	AAA	GAA	CGT	2910
73	V	C	L	L	Y	D	Q	W	V	L	S	P	H	K	K	E	R	92
2911	CAC	CTG	GGT	AAC	CTG	GTA	ATC	ACC	TGG	GGC	GCC	CAG	ACT	TTC	AAA	CAC	CAA	2970
93	H	L	G	N	L	V	I	T	W	G	A	Q	T	F	K	H	Q	112
2971	AAA	CTG	GCT	AAC	CTG	TTC	ATC	GTT	AAC	AAC	AAA	ACC	ATC	CCG	AAC	AAC	CTG	3030
113	K	L	A	N	L	F	I	V	N	K	K	T	I	P	N	N	L	132
3031	AAC	TAC	CTG	ACC	CCG	ATG	TCT	CTG	GCA	TAC	TGG	TTC	ATG	GAT	GAT	GGT	GGT	3090
133	N	Y	L	T	P	M	S	L	A	Y	W	F	M	D	G	G	K	152
3091	TAC	AAC	AAA	AAC	TCT	ACC	AAC	AAA	TCG	ATC	GTA	CTG	AAC	ACC	CAG	TCT	TTC	3150
153	Y	N	K	N	S	T	N	K	S	I	V	L	N	T	Q	S	F	172
3151	GAA	GTA	GAA	TAC	CTG	GTT	AAG	GGT	CTG	CGT	AAC	AAA	TTC	CAA	CTG	AAC	TGT	3210
173	E	V	E	Y	L	V	K	G	L	R	N	K	F	Q	L	N	C	192
3211	ATC	AAC	AAA	AAC	AAA	CCG	ATC	ATC	TAC	ATC	GAT	TCT	ATG	TCT	TAC	CTG	ATC	3270
193	I	N	K	N	K	P	I	I	Y	I	D	S	M	S	Y	L	I	212
3271	CTG	ATC	AAA	CCG	TAC	CTG	ATC	CCG	CAG	ATG	ATG	TAC	AAA	CTG	CCG	AAC	ACT	3330
213	L	I	K	P	Y	L	I	P	Q	M	Y	K	L	P	N	T	I	232
3331	GAA	ACT	TTC	CTG	AAA	TAA	SalI											3404
233	E	T	F	L	K		pstI											238
							TAAGTCGACCTGCAGCCCAAGCTTGGCACTGGCCGCTGTTTACACGTCGTGACT											

FIG. 4B



		-2	-1	1				5					10							
		M	H	M	K	N	I	K	K	N	Q	V	M	N	L	G	P	N	S	
K	L	L	20 K	E	Y	K	S	Q	L	I	E	L	30 N	I	E	Q	F	E	A	
G	I	G	40 L	I	L	G	D	A	Y	I	R	S	50 R	D	E	G	K	T	Y	
C	M	Q	60 F	E	W	K	N	K	A	Y	M	D	70 H	V	C	L	L	Y	C	
Q	W	Y	80 L	S	P	P	H	K	K	E	R	Y	90 N	H	L	G	N	L	Y	
I	T	W	100 G	A	Q	T	F	K	H	Q	A	F	110 N	K	L	A	N	L	F	
I	V	N	120 N	K	K	I	I	P	N	N	L	V	130 E	N	Y	L	T	P	M	
G	L	A	140 Y	W	P	M	D	D	G	G	K	W	150 D	Y	N	K	N	S	I	
N	K	S	160 I	V	L	N	T	Q	S	F	T	F	170 E	E	V	E	Y	L	V	
K	G	L	180 R	N	K	F	Q	L	N	C	Y	V	190 K	I	N	K	N	K	P	
I	I	Y	200 I	D	S	M	S	Y	L	I	F	Y	210 N	L	I	K	P	Y	L	
I	P	Q	220 M	M	Y	K	L	P	N	T	I	S	230 S	E	T	F	L	K	*	

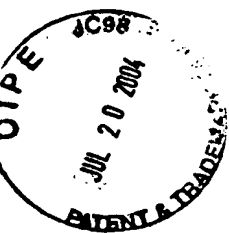
Positions that can be changed without affecting enzyme activity (demonstrated)  
positions -1 and -2 are not natural. The two amino acids are added due to cloning strategies

positions 1 to 10: can be deleted  
position 36: G is tolerated  
position 40: M or V are tolerated  
position 41: S or N are tolerated  
position 43: A is tolerated  
position 46: V or N are tolerated  
position 91: A is tolerated  
positions 123 and 156: L are tolerated  
position 223: A and S are tolerated

Changes that affect enzyme activity (demonstrated)

position 19: L to S  
position 38: I to S or N  
position 39: G to D or R  
position 40: L to Q  
position 42: L to R  
position 44: D to E G or H  
position 45: A to E or D  
position 46: Y to D  
position 47: I to R or N  
position 80: L to S  
position 144: D to E  
position 145: D to E  
position 146: G to E  
position 147: G to S

FIG. 5



Group I Intron Encoded Endonucleases and Related Endonucleases

ENDONUCLEASE		RECOGNITION SEQUENCE	CLEAVAGE SITE	▽ INTRON SITE
TWO DODECAPEPTIDE FAMILY (OR 4 BP CUTTERS)	I-Sce I (Saccharomyces mitochondria)	CGCTAGGGATAACAGGGTAATATAGC GCGATCCCATTATTGTCCATTATATCG		
	I-Sce IV (Saccharomyces mitochondria)	TTCTCATGATTAGCTCTAATCCATGG AAGAGTACTAATCGAGATTAGGTACC		
	I-Sce II (Saccharomyces mitochondria)	CTTTGGTCAATCAGAAAGTATATATTT GAAACCAGTAGGTCTTCATATATAAA		
	I-Ceu I (Chlamydomonas chloroplast)	TAA CGGTCCIAAGGTAGCGAAATTCA ATTGCCAGGATTCCATCGCTTTAAGT		
	I-Ppo I (Physarum nucleus)	TGACTCTCTIAAGGTAGCCAAATGCC ACTGAGAGAATTCCATCGGTTTACGG		
	I-Sce III (Saccharomyces mitochondria)	GGAGGTTTTGGTAACTATTTATTACC CCTCCAAAACCATTGATAAATAATGG		
	I-Cre I (Chlamydomonas chloroplast)	GGGTTCAAACGTCGTGAGACAGTTT CCCAAGTTTTGCAGCACTCTGTCAA		
	Endo. Sce I(RF3) (Saccharomyces mitochondria) (Non intronic)	GATGCTGTAGGCATAGGCTTGTTAT CTACGACATCCGTATCCGAACCAATA		
	HO (Saccharomyces nucleus) (Non intronic)	CTTTCCGCAACAGTATAATTTTATAA GAAAGGCGTTGTCAATTAAAAATATT		
	I-Csm I (Chlamydomonas mitochondria) (Putative endonuclease)	ACCATGGGGTCAAATGTCTTTCTGGG TGGTACCCAGTTTACAGAAAGACCC		
	I-Pan I (Podospira mitochondria) (Putative endonuclease)	GTGCCTGAATGATATTTATTACCTTT CACGGACTTACTATAAATAATGGAAA		
	(Bacteriophage T4)			
OTHER STRUCTURAL FAMILIES	I Tev I	CAACGCTCAGTAGATGTTTTCTTGGGTCTACCGTTTAAT GTTCGAGTCATCTACAAAAGAACCCAGATGGCAAATTA		
	I Tev II	CAAGCTTATGAGTATGAAGTGAACACGTTATT GTTCGAATACTCATACTTCACTTGTGCAATAA		
	I Tev III	GCTATTGTTTTTATGATCTTTTGGGTGTAGCTTTAA CGATAAGCAAAAATACATAGAAAACGCACATCGAAATT		

FIG. 6

OIPF  
JUL 20 2004  
PATENT & TRADEMARK

EXPRESSION VECTORS

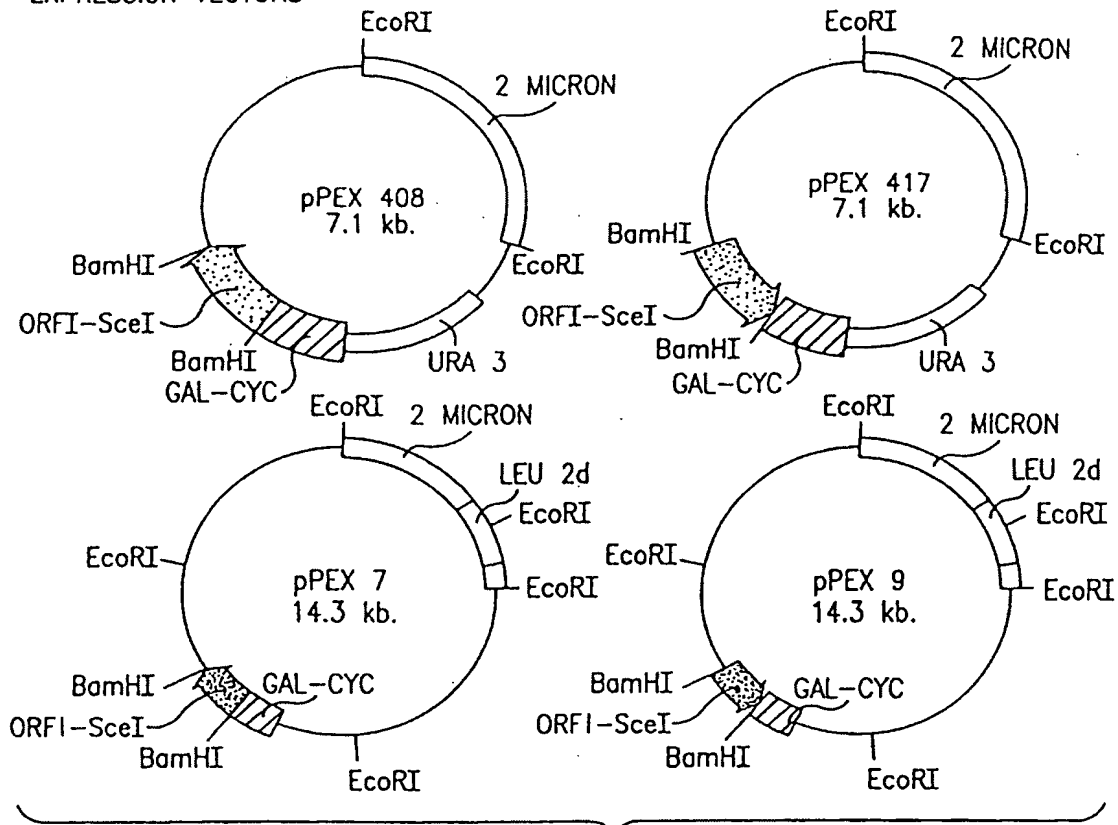


FIG. 7

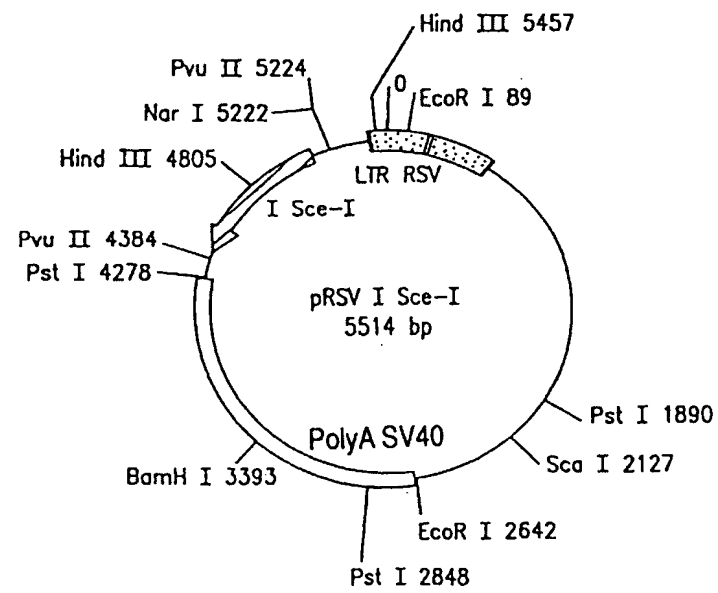


FIG. 8



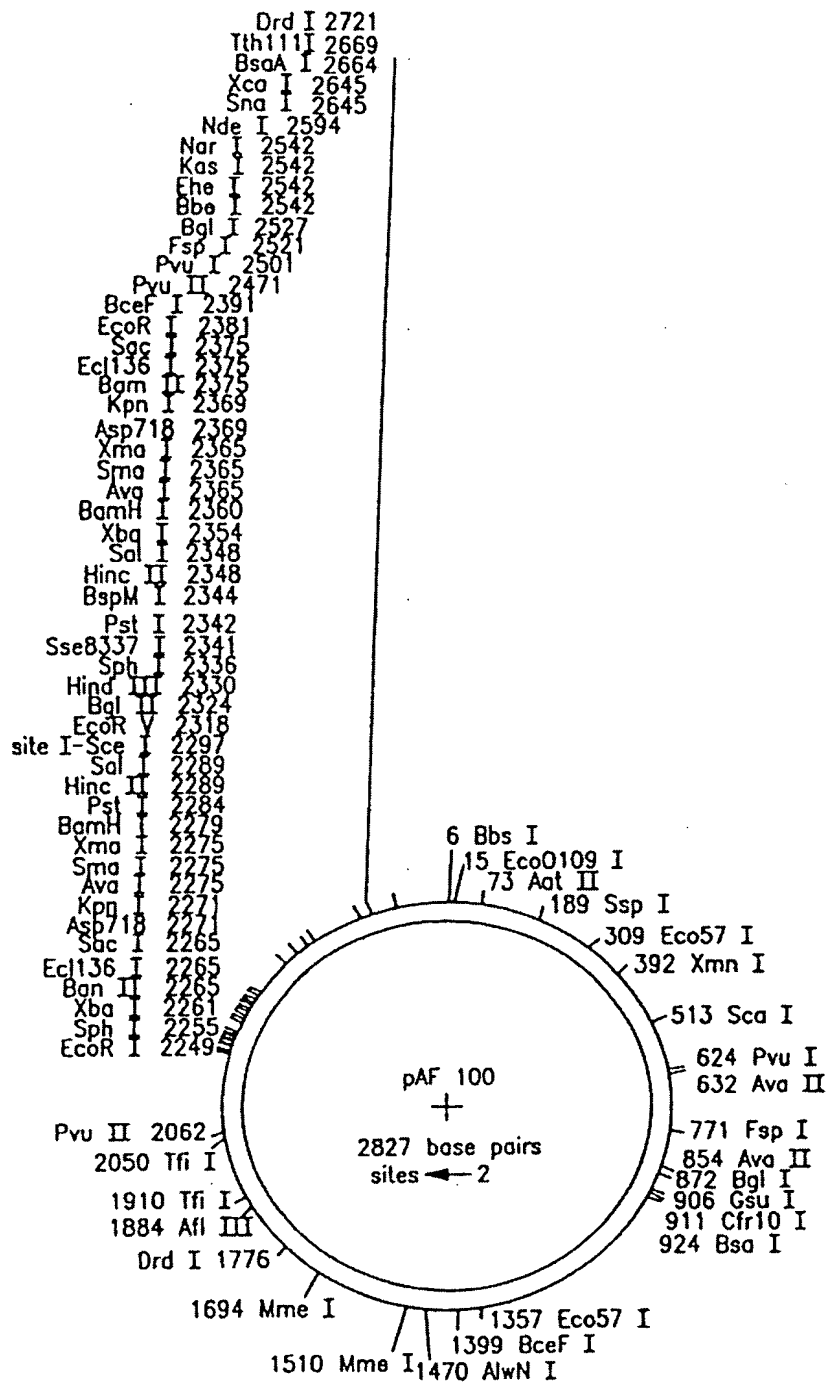


FIG. 9

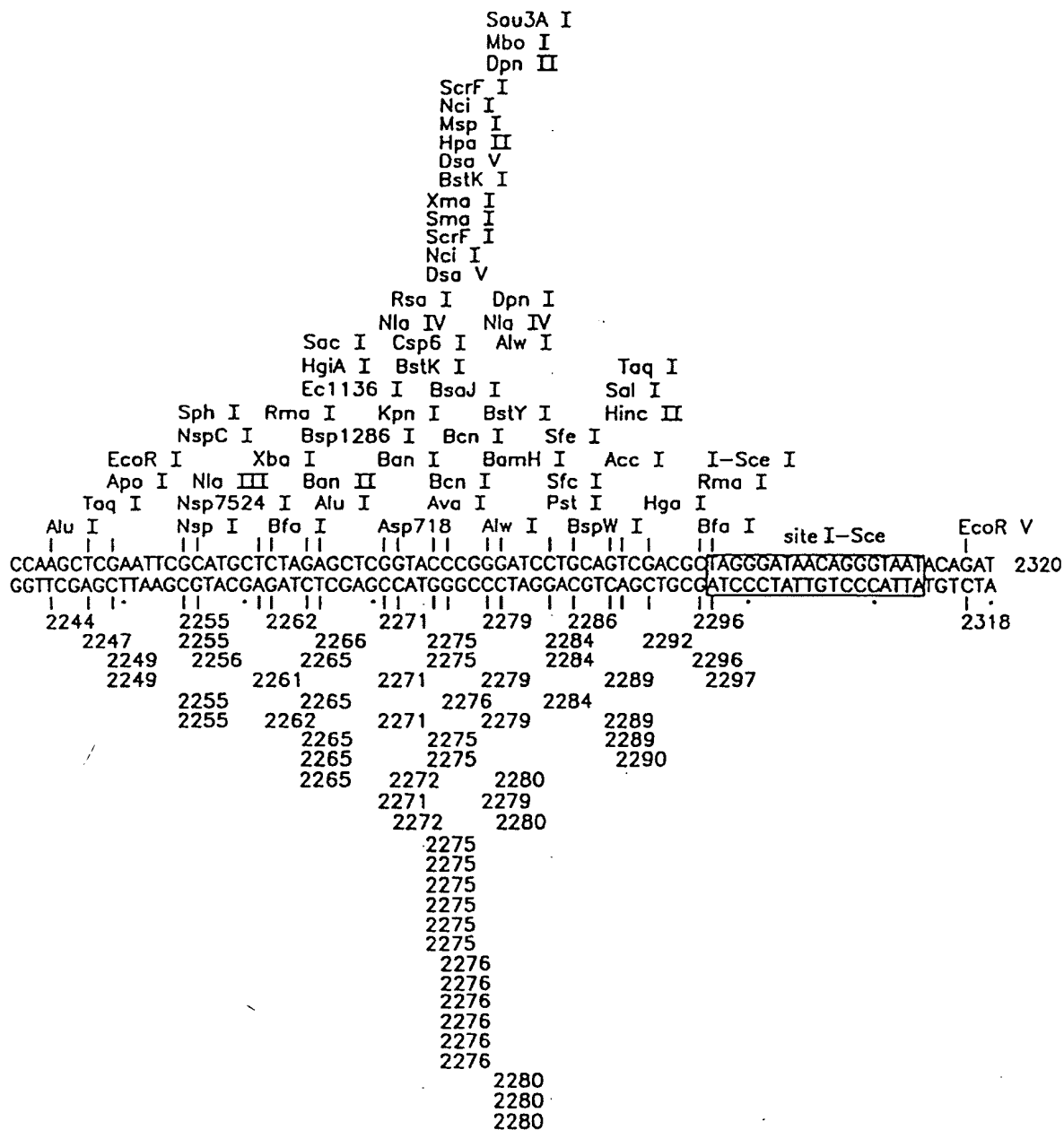
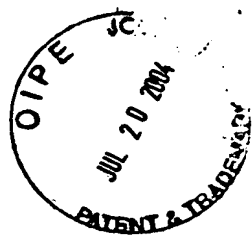


FIG. 10A

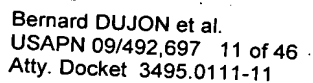
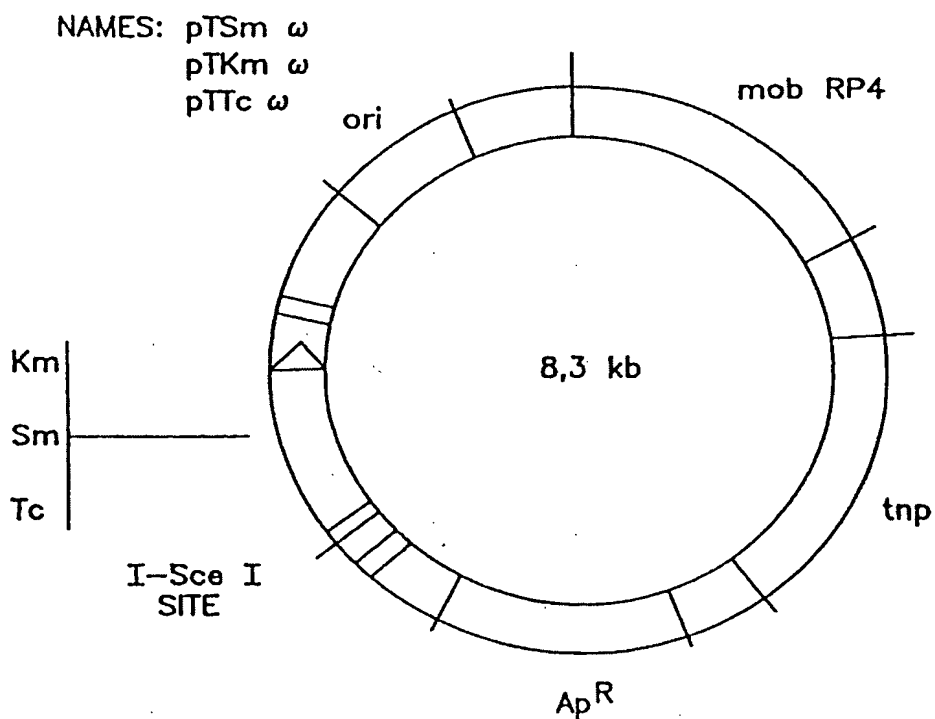
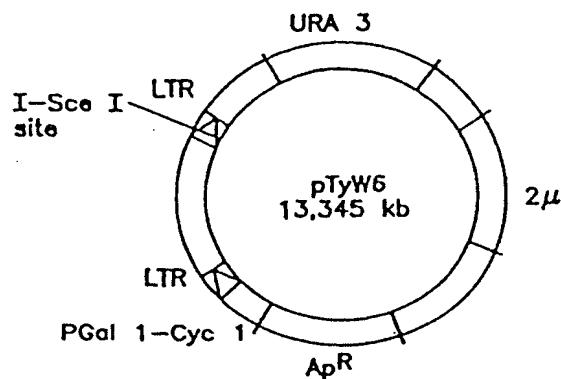


FIG. 10B



Construction: pGP 704 from De Lorenzo, with transposase gene and insertion of the linker[I-SceI] in NotI unique site

FIG. 11



Construction: pD 123, from J.D. Boeke  
with insertion of a linker[I-SceI-NotI] in BamHI

FIG. 12

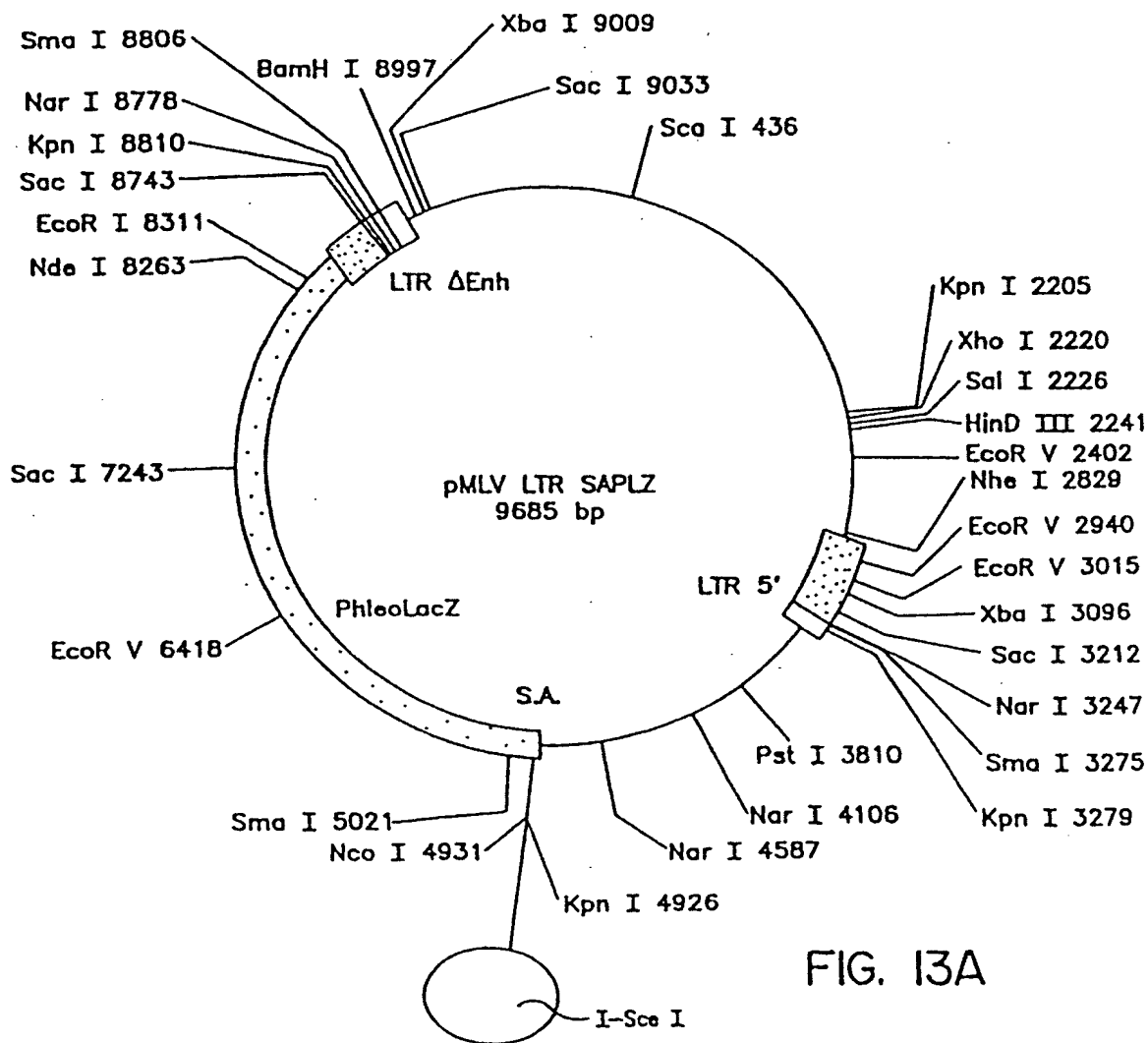


FIG. 13A

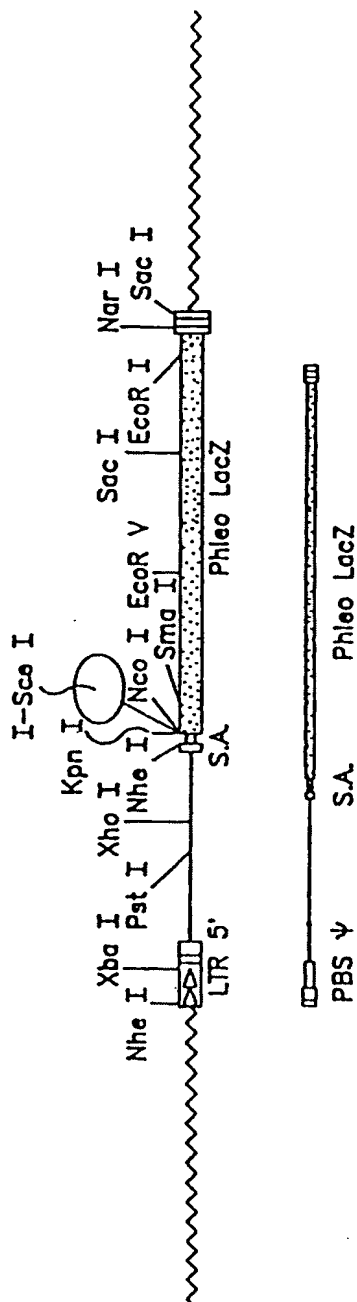
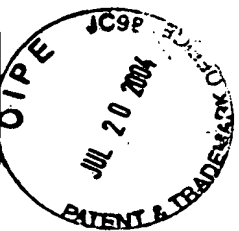


FIG. 13B

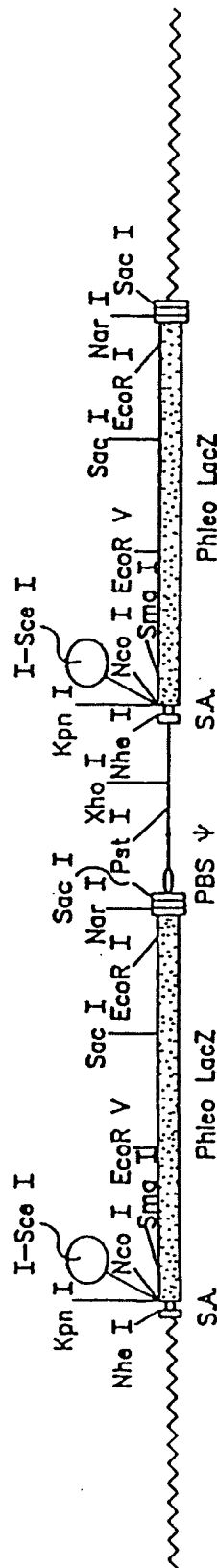


FIG. 13C



CONTROL D304 G41 M57 CONTROL  
A302 E40 H81 T62

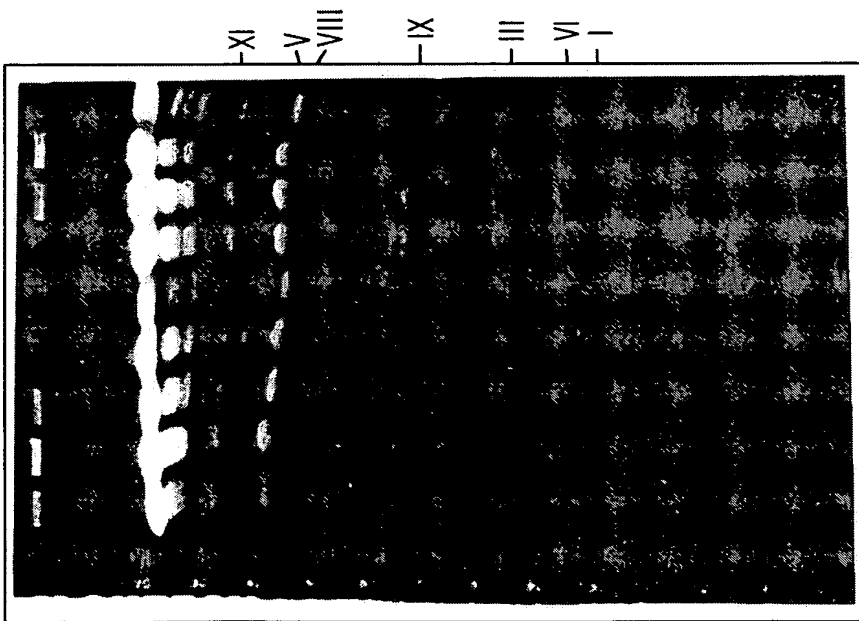
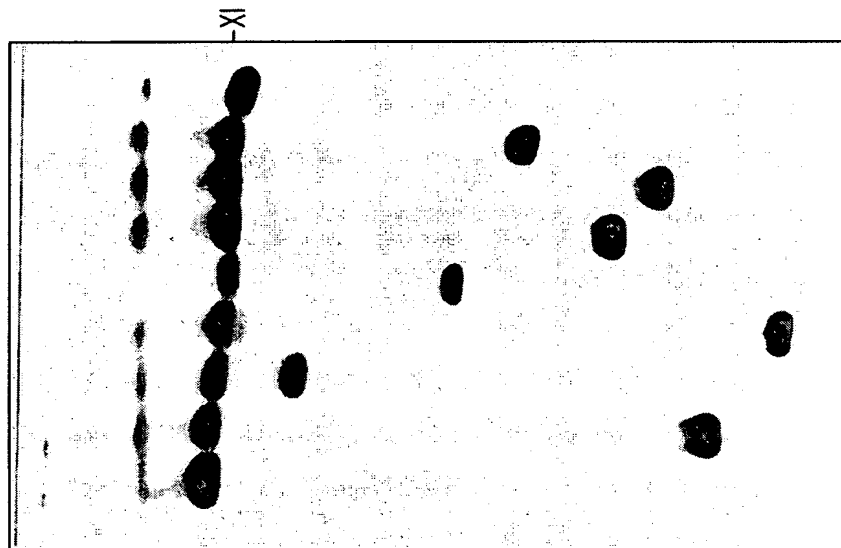


FIG. 14A

CONTROL A304 G41 M57 CONTROL  
A302 E40 H81 T62



LEFT END PROBE  
COSMID pUKG 040

FIG. 14B

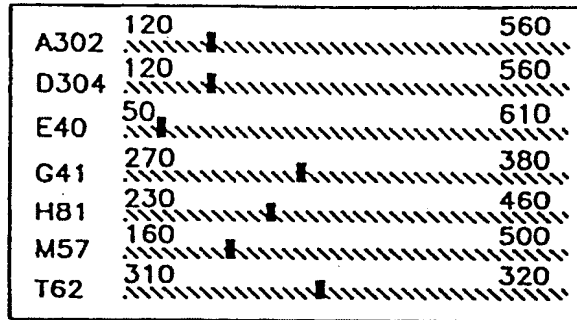
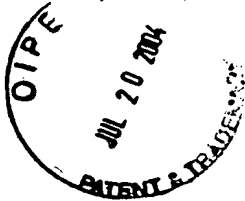


FIG. 15A

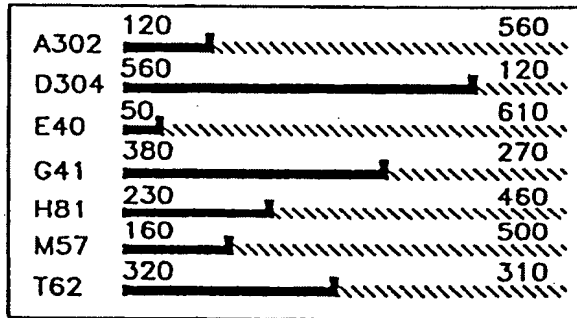


FIG. 15B

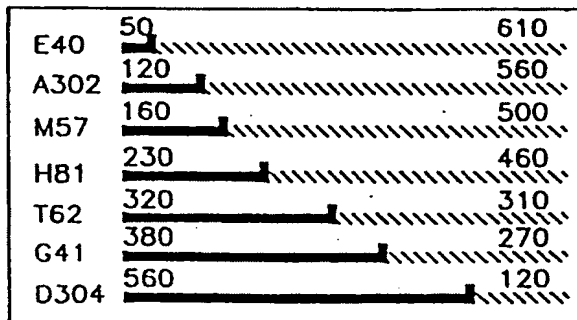


FIG. 15C





FIG. 15D

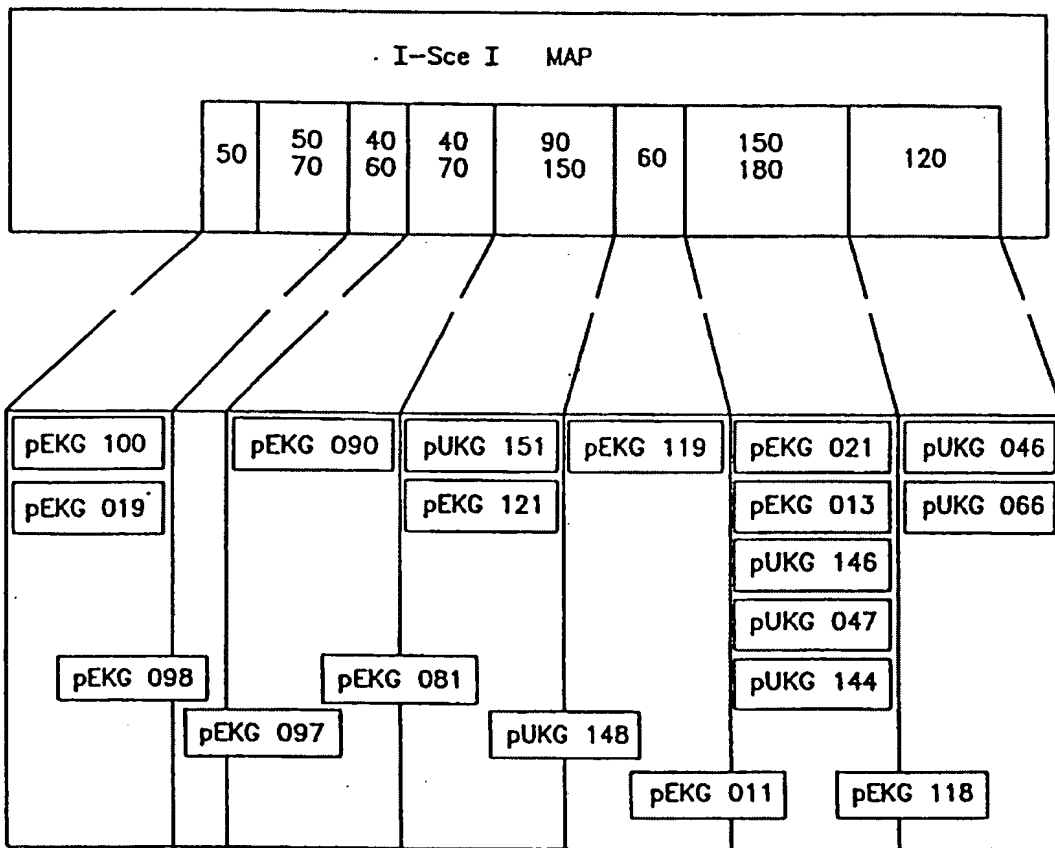


FIG. 15E

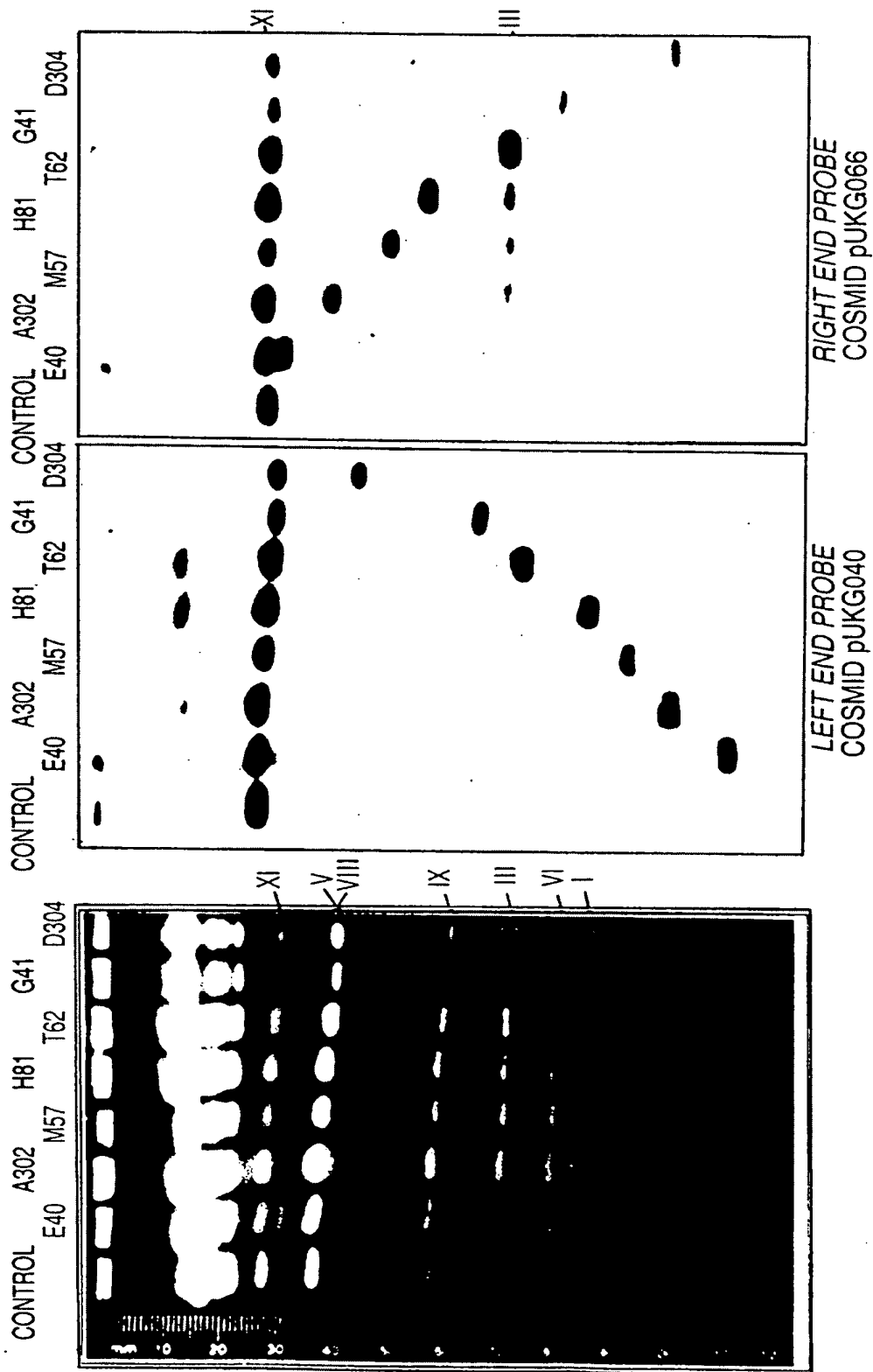
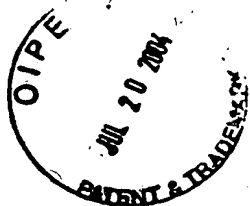


FIG. 16C

FIG. 16B

FIG. 16A

This image is a high-contrast, black and white scan of a document page. The top half of the page is mostly white, with a few small, dark, irregular spots near the center. The bottom half of the page is predominantly black, with a dense pattern of small, bright, rectangular spots arranged in vertical columns, resembling a film strip or a heavily degraded document page. The overall appearance is that of a low-quality, high-contrast scan of a document page.

FIG. 17B

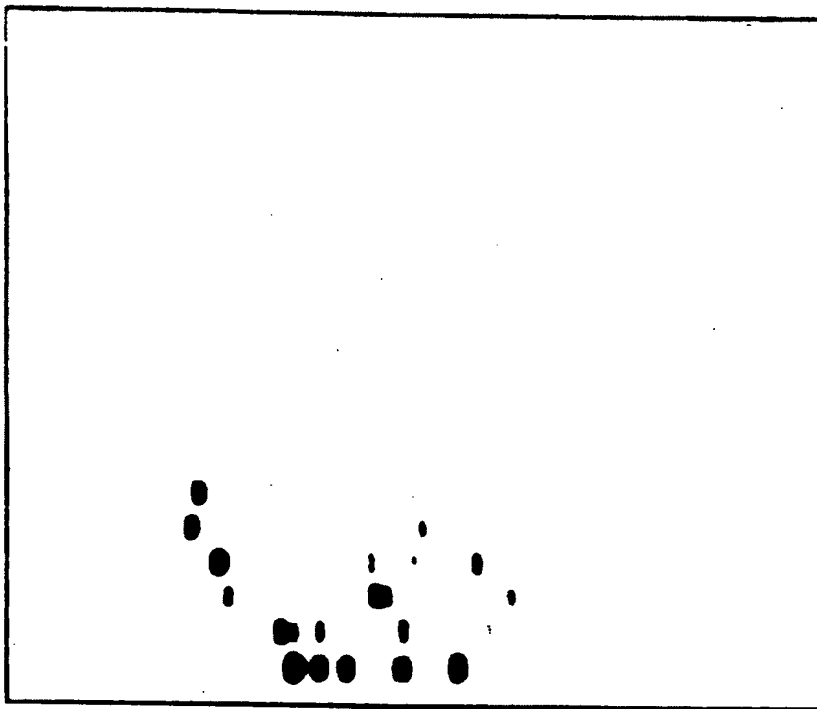
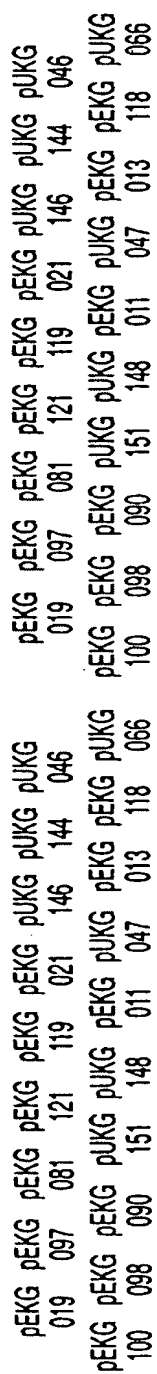
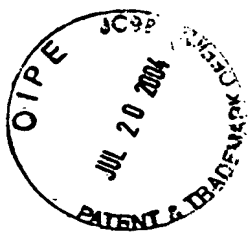


FIG. 17C

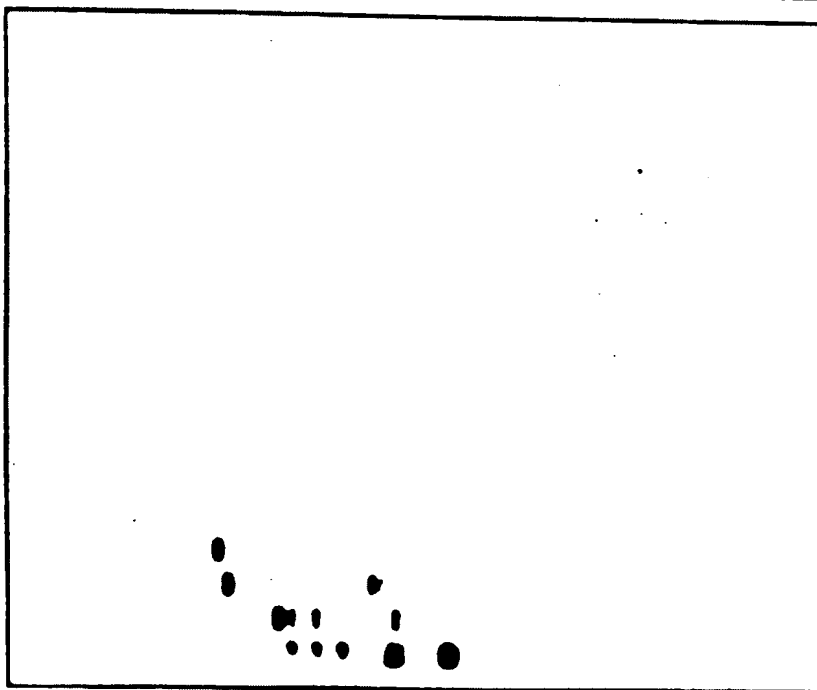


FIG. 17D

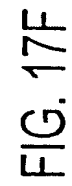
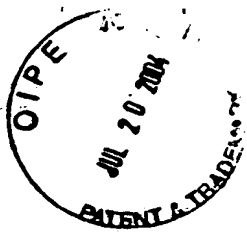


FIG. 17E



pEKG	pEKG	pEKG	pEKG	pEKG	pEKG	pEKG	pEKG
019	097	081	121	119	021	146	046
pEKG	pEKG	pEKG	pUJG	pUJG	pEKG	pEKG	pUJG
100	098	090	151	148	011	047	013
pEKG	pEKG	pEKG	pUJG	pUJG	pEKG	pEKG	pUJG
019	097	081	121	119	021	146	046
pEKG	pEKG	pEKG	pUJG	pUJG	pEKG	pEKG	pUJG
100	098	090	151	148	011	047	013
pEKG	pEKG	pEKG	pUJG	pUJG	pEKG	pEKG	pUJG
019	097	081	121	119	021	146	046

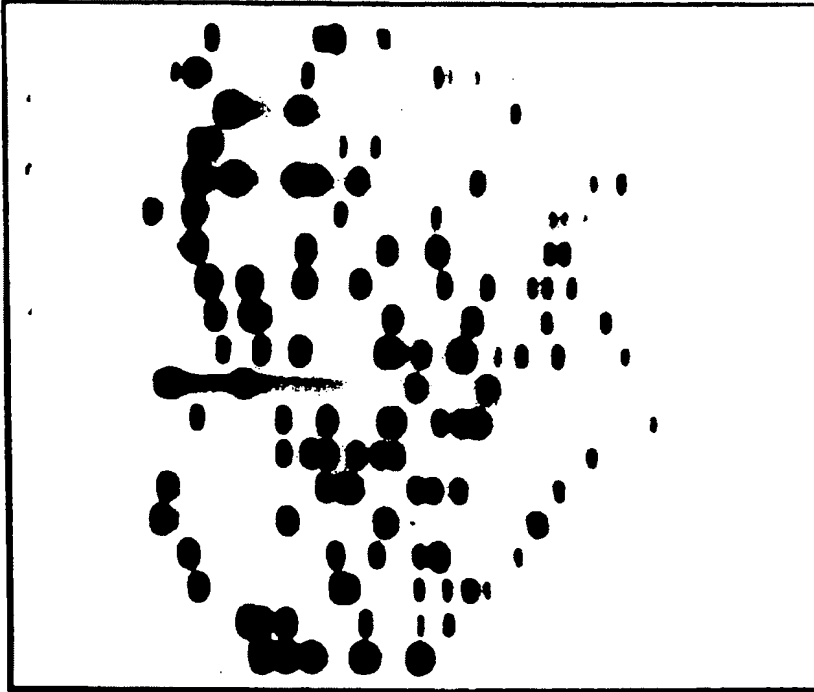


FIG. 17H

pEKG pEKG pEKG pEKG pEKG pEKG pEKG  
019 097 081 121 119 021 146 144 046

pEKG pEKG pEKG pUJG pUJG pUJG pUJG  
100 098 090 151 148 011 047 013 118 066

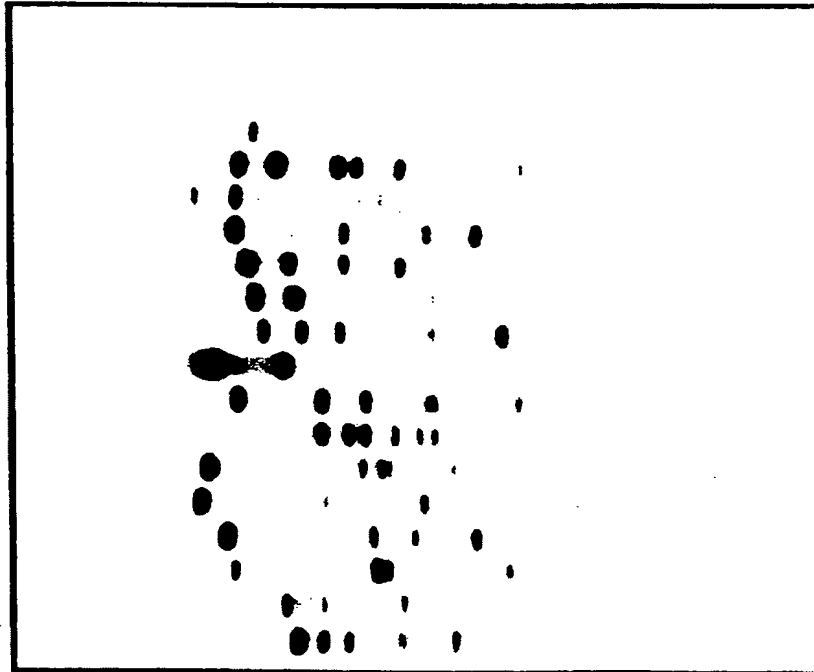


FIG. 17G

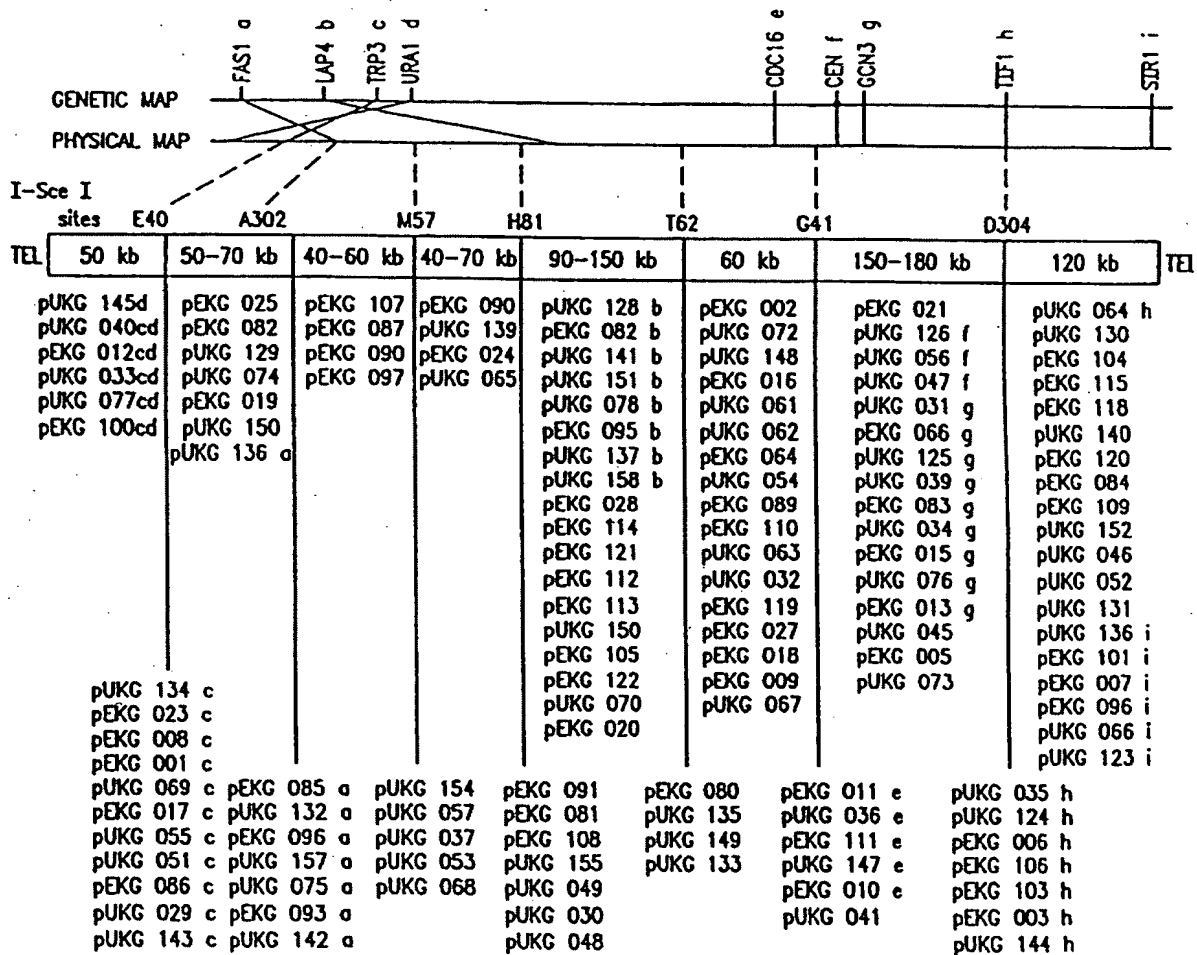
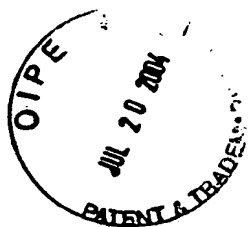


FIG. 18

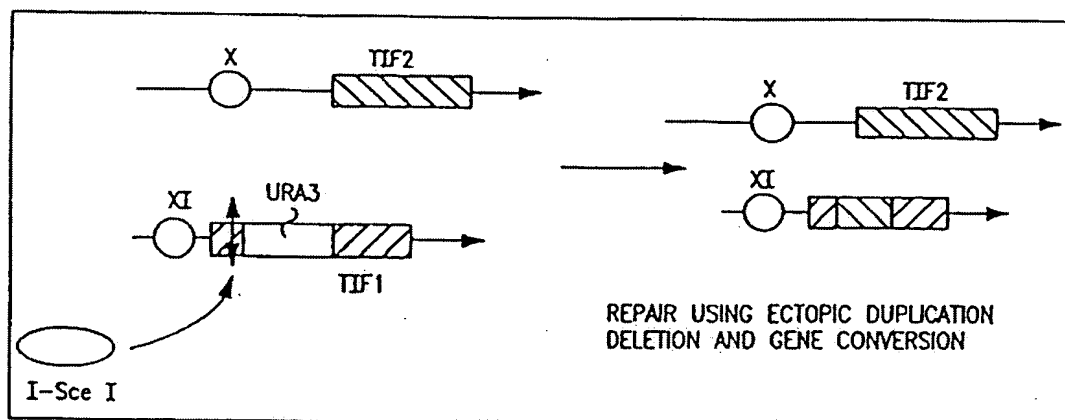
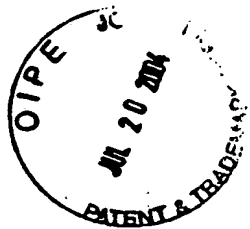


FIG. 19A

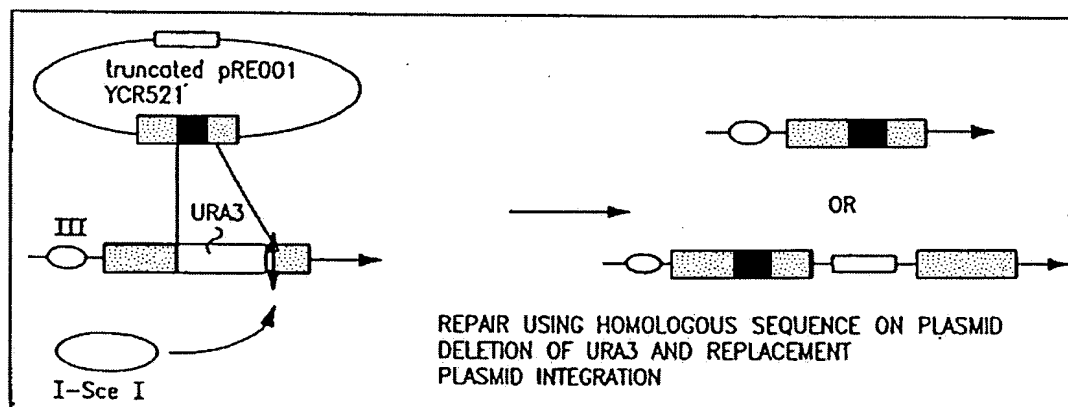
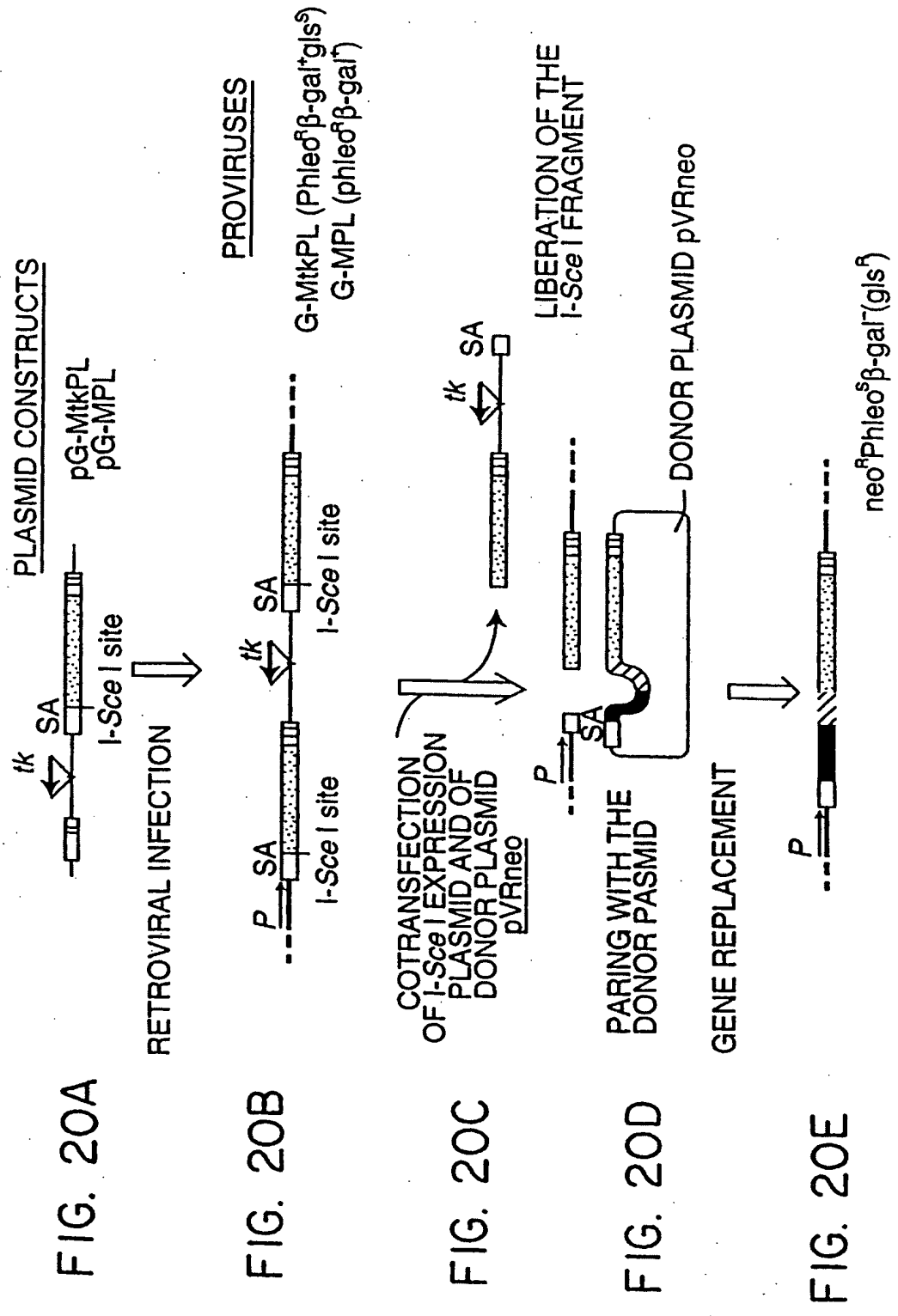


FIG. 19B





LTR PhleoLacZ NEO POLY A GENOMIC DNA tk THYMIDINE KINASE

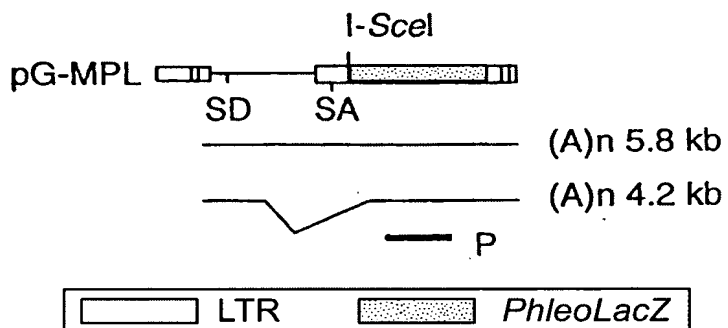


FIG. 2IA

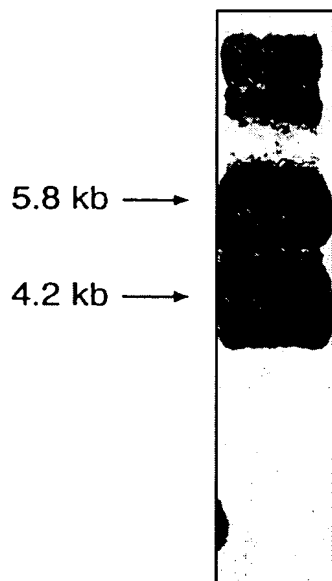


FIG. 21B

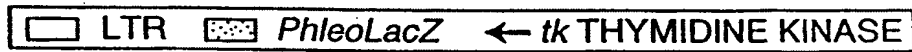
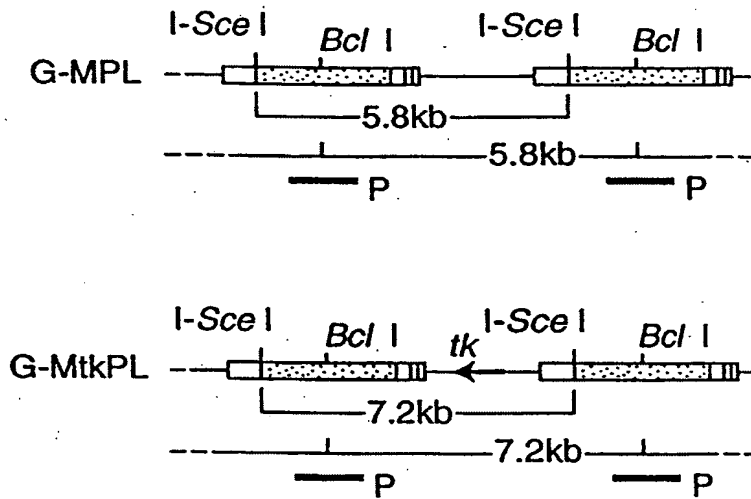


FIG. 22A

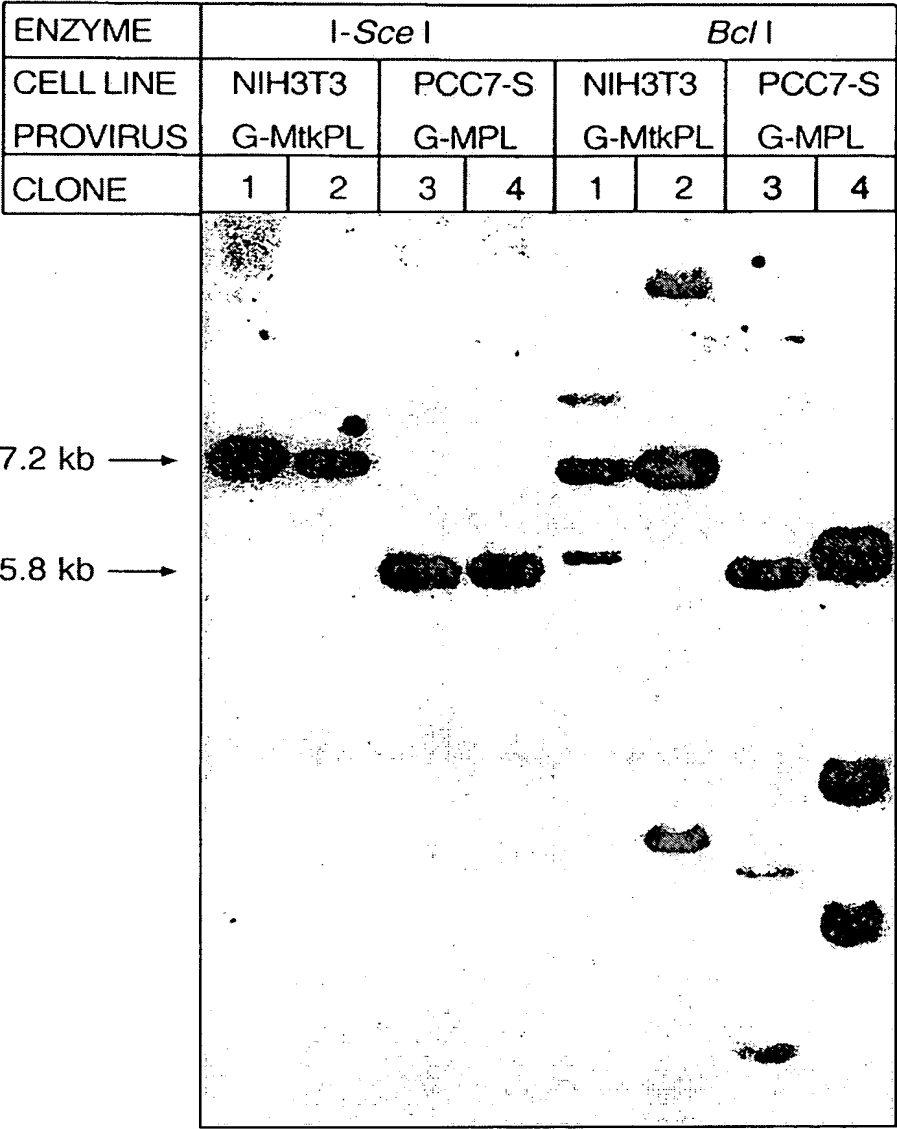


FIG. 22B

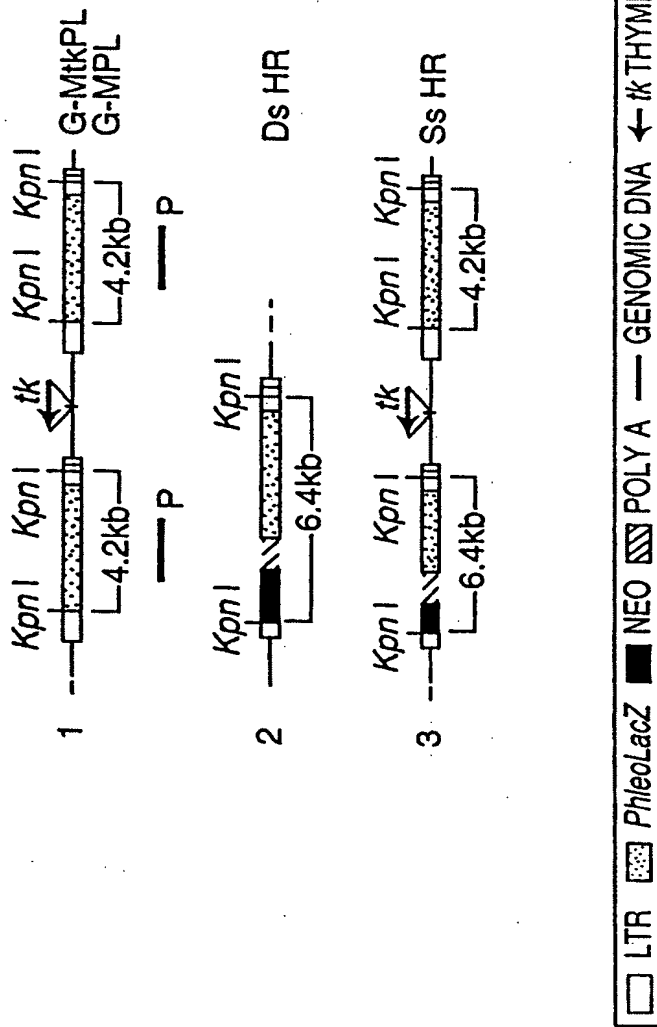


FIG. 23A

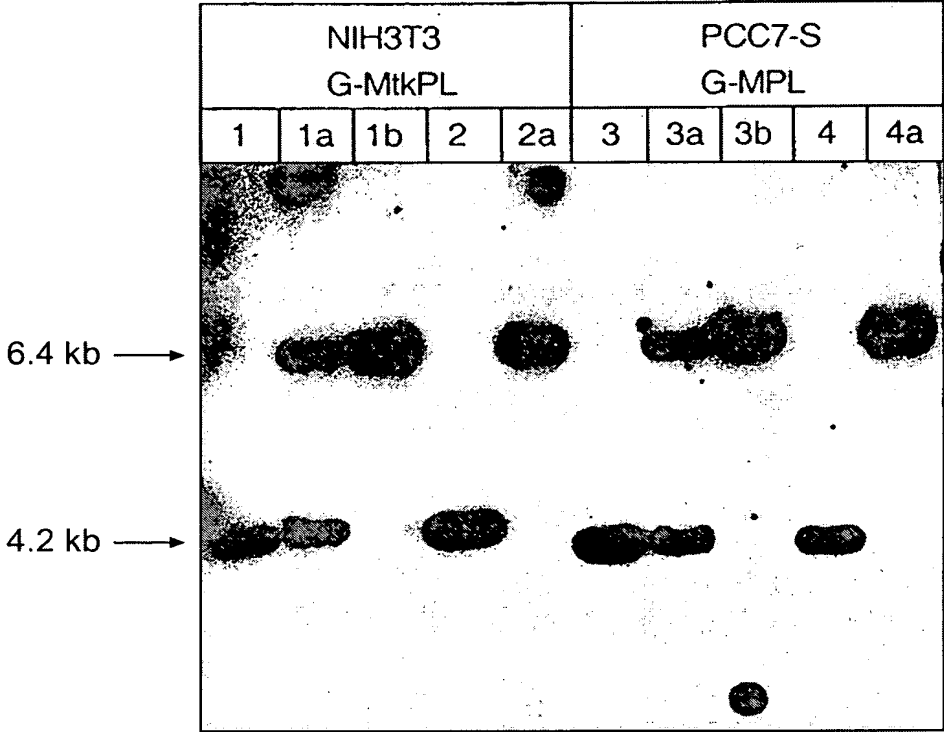


FIG. 23B



PCC7-S / G-MPL, CLONE 3

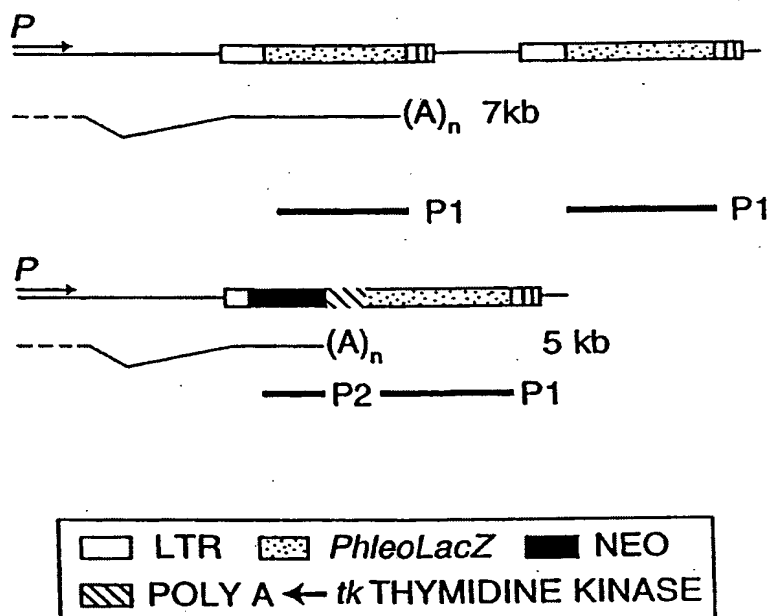


FIG. 24A



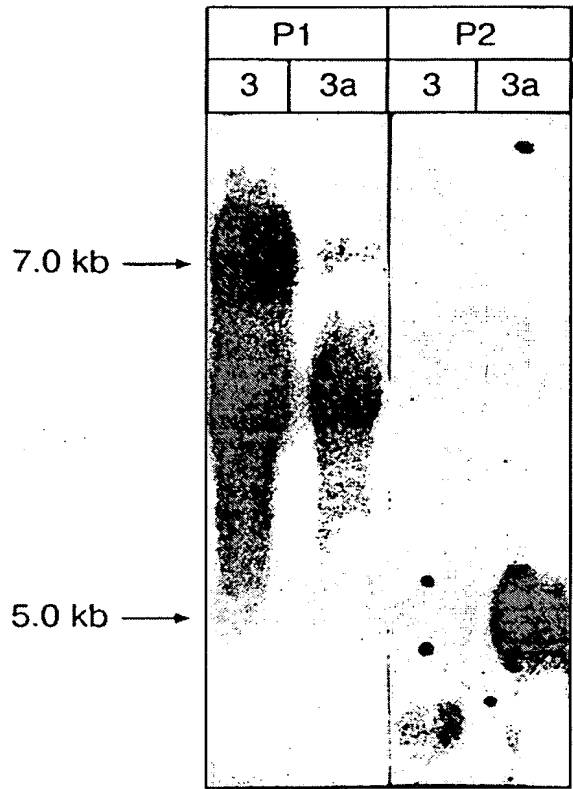
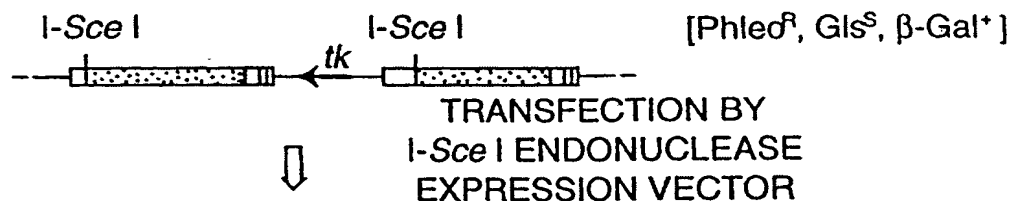


FIG. 24B



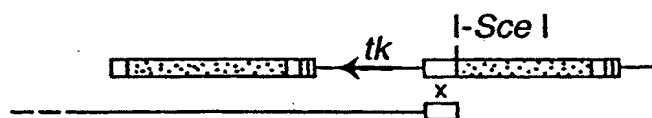
A. CHROMOSOMAL DNA  
CONTAINING PROVIRUS

PHENOTYPES

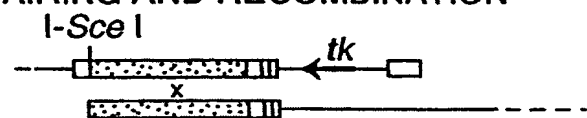


B. INTRA-CHROMOSOMAL  
RECOMBINATIONS EVENTS

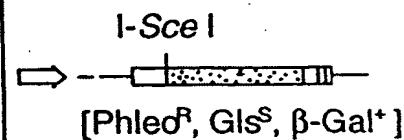
1. THE LEFT I-Sce I IS CUT.  
PAIRING AND RECOMBINATION



2. THE RIGHT I-Sce I IS CUT.  
PAIRING AND RECOMBINATION



3. BOTH I-Sce I SITES ARE CUT.  
RELIGATION BY END-JOINING



C. INTER-CHROMOSOMAL RECOMBINATION EVENT  
BOTH I-Sce I SITES ARE CUT. GAP REPAIR USING INTACT  
CHROMOSOME SEQUENCES

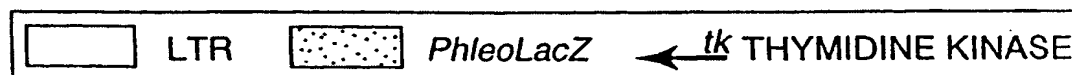
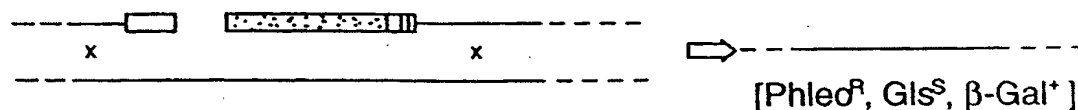


FIG. 25



A. PARENTAL DNA, G-MtkPL

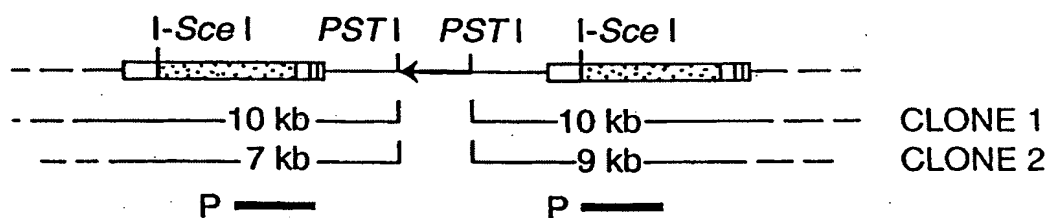


FIG. 26A

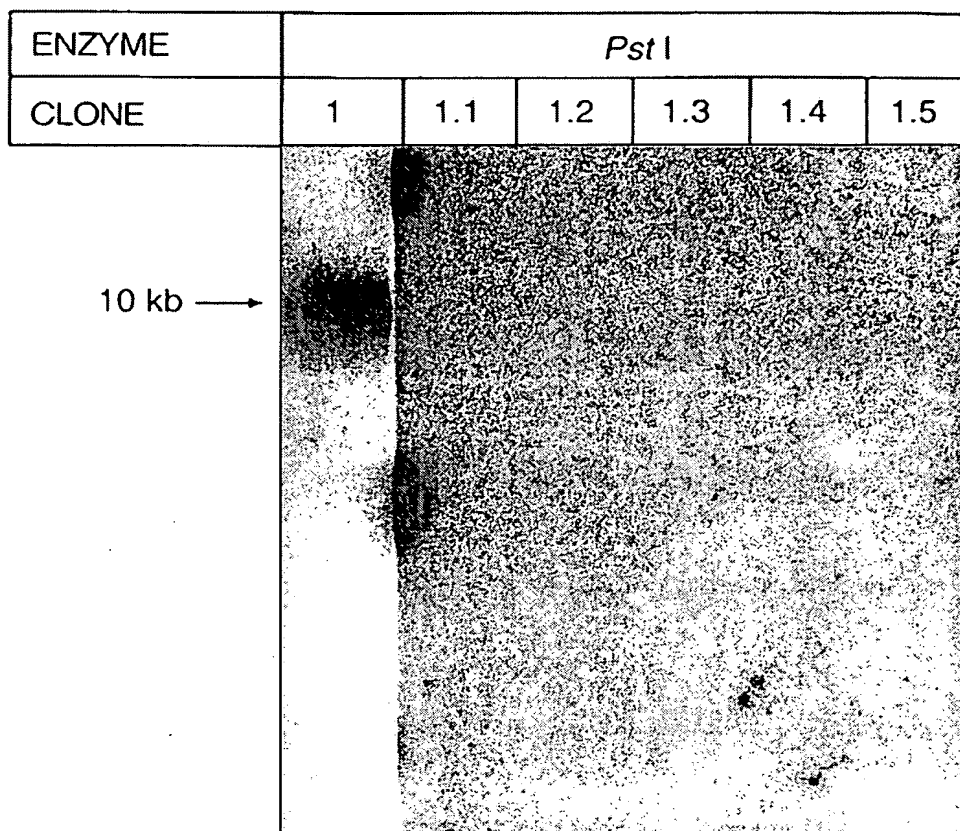
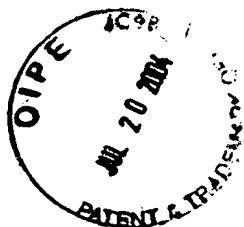


FIG. 26B

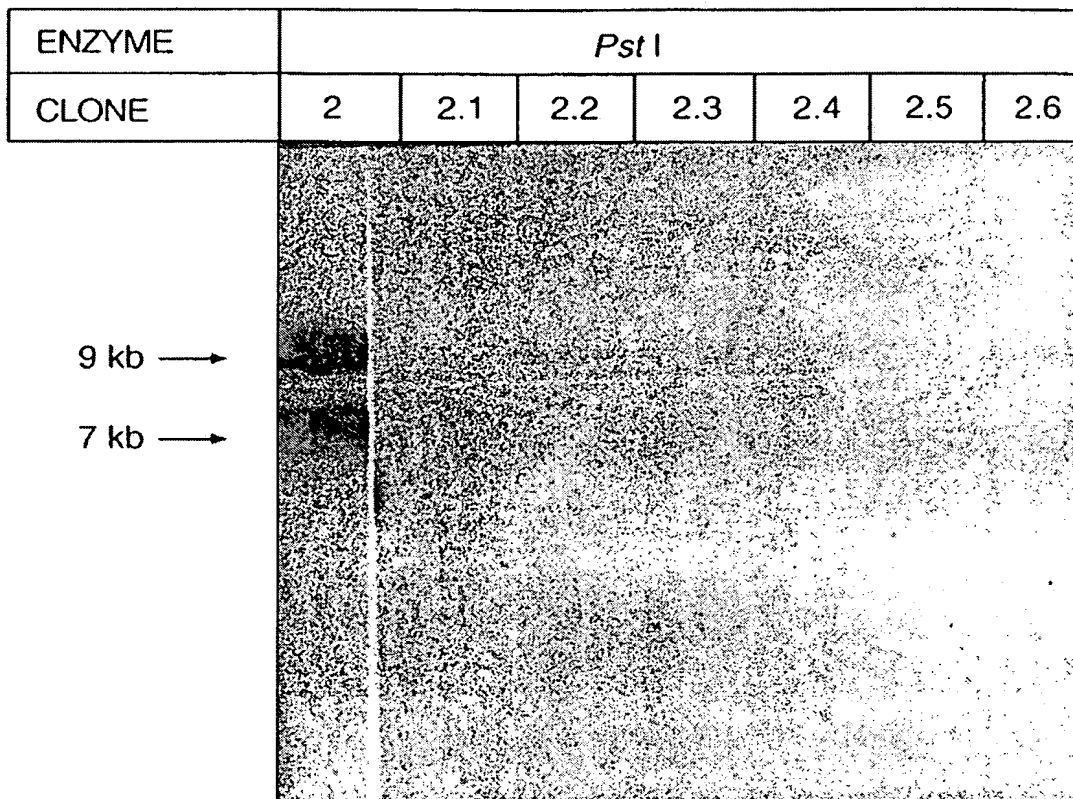
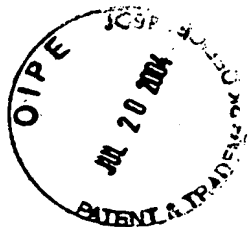
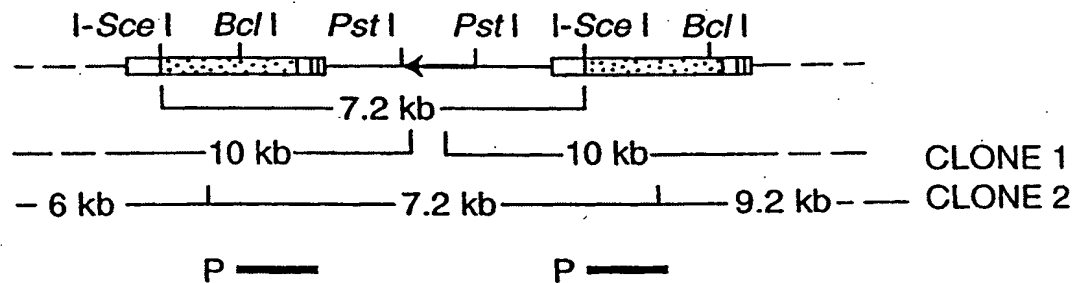


FIG. 26C



### 1. PARENTAL DNA, G-MtkPL



### 2. INTRA-MOLECULAR RECOMBINATION EVENT

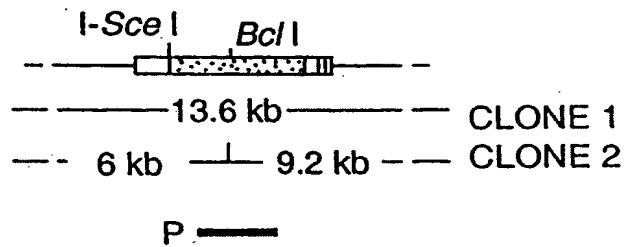


FIG. 27A

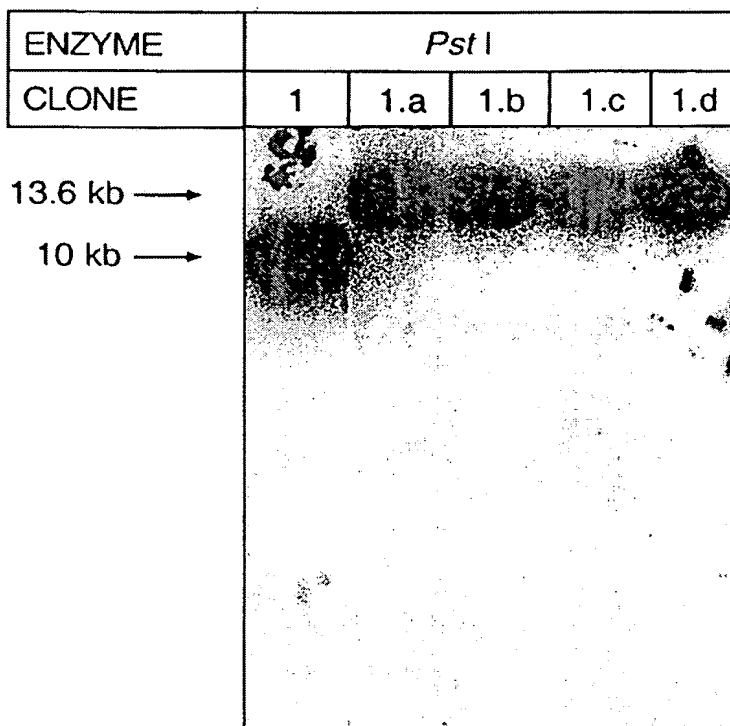
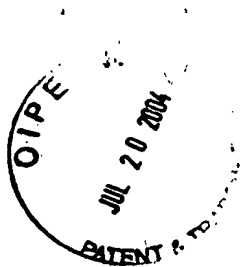


FIG. 27B



9

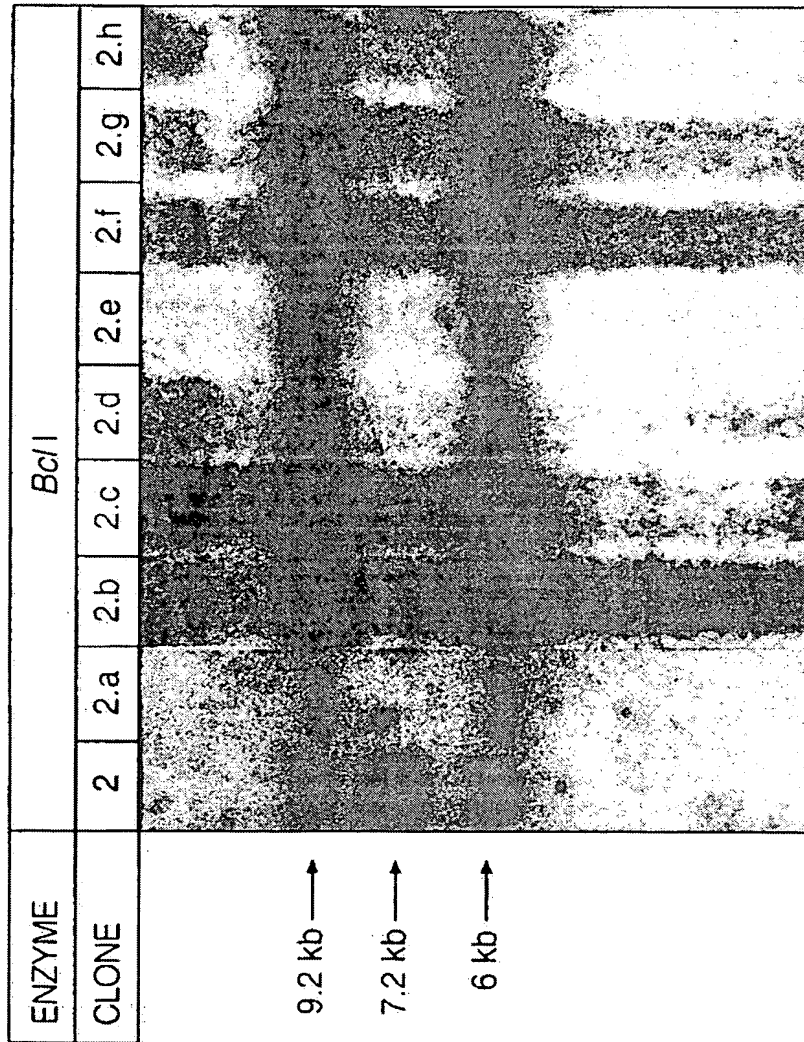


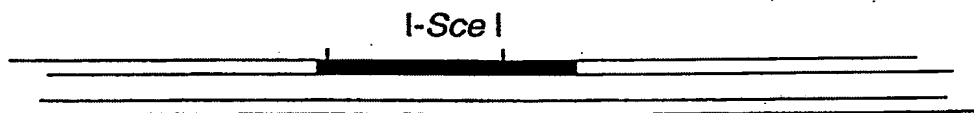
FIG. 27C



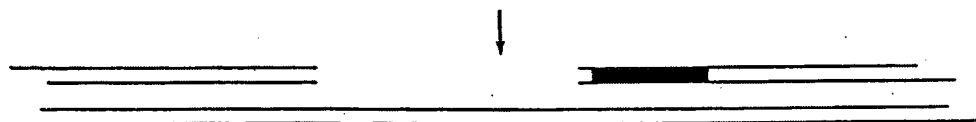


## LOSS OF HETEROZYGOSITY

INTEGRATION OF ARTIFICIAL SITE OR  
PRESENCE OF SPECIFIC SITE



EXPRESSION OF I-Sce I AND SPECIFIC CLEAVAGE



REPAIR OF THE DSB WITH THE OTHER CHROMATID

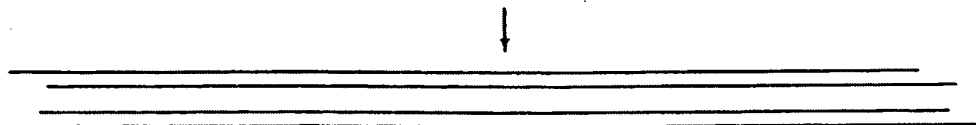
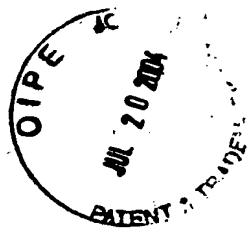


FIG. 28



### CONDITIONAL ACTIVATION (TANDEM REPEAT)

#### INTEGRATION OF ARTIFICIAL SITE BETWEEN TANDEM REPEATS



GENE X INACTIVE

#### EXPRESSION OF I-Sce I AND SPECIFIC CLEAVAGE

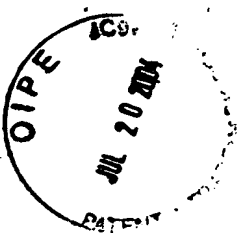


#### REPAIR OF THE DSB BY SINGLE STRAND ANNEALING



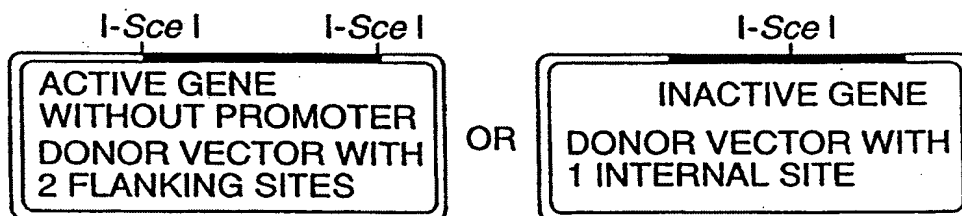
GENE X ACTIVE

FIG. 29

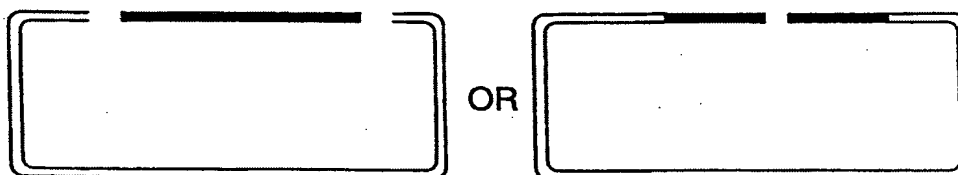


## ONE STEP REARRANGEMENT

INTEGRATION OF ARTIFICIAL SITE OR  
PRESENCE OF SPECIFIC SITE



EXPRESSION OF I-Sce I ENZYME  
AND  
SPECIFIC CLEAVAGE OF THE DONOR PLASMID



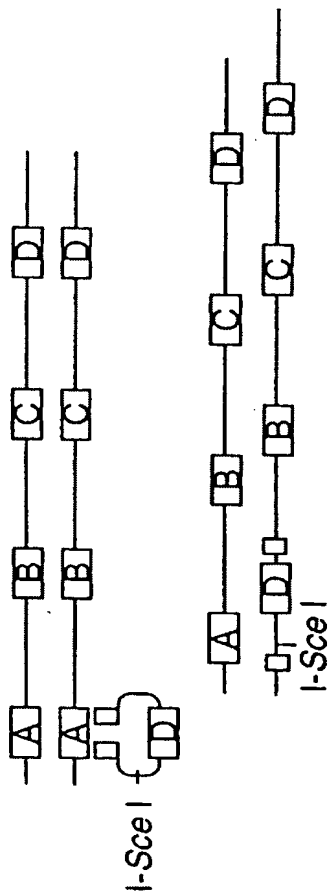
RECOMBINATION BETWEEN THE CHROMOSOME AND PLASMID



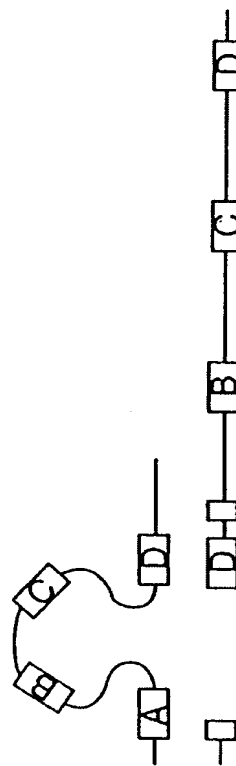
FIG. 30

# DUPLICATION OF A LOCUS

## 1. INSERTION OF I-Sce I SITE BY CLASSICAL GENE REPLACEMENT



## 2. SPECIFIC CLEAVAGE BY I-Sce I ENZYME AND REPAIR OF THE BREAK BY HOMOLOGOUS SEQUENCES



## 3. DUPLICATION OF THE TOTALITY OF THE LOCUS

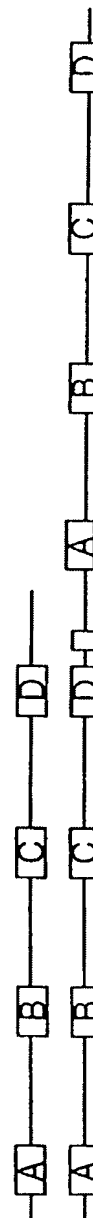
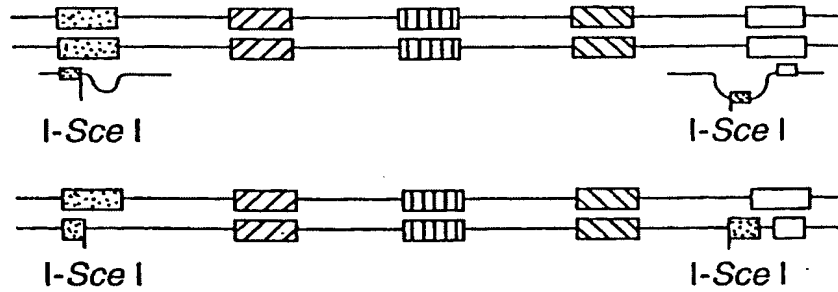


FIG. 3I

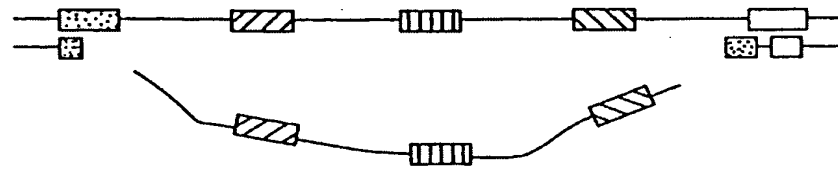


## DELETION OF A LOCUS

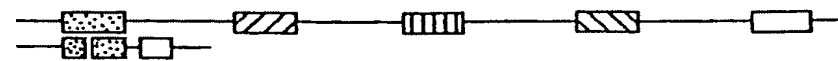
### 1 INSERTION OF TWO I-Sce I SITES FLANKING THE LOCUS



### 2 EXPRESSION OF THE ENZYME AND CLEAVAGE



### 3 RECOMBINATION BETWEEN THE TWO ENDS



### 4 DELETION OF THE LOCUS

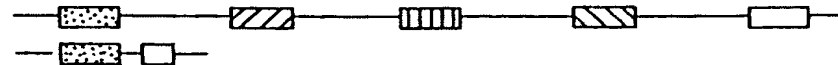


FIG. 32

OIPF  
JUL 20 2004  
PATENT & TRADEMARK OFFICE

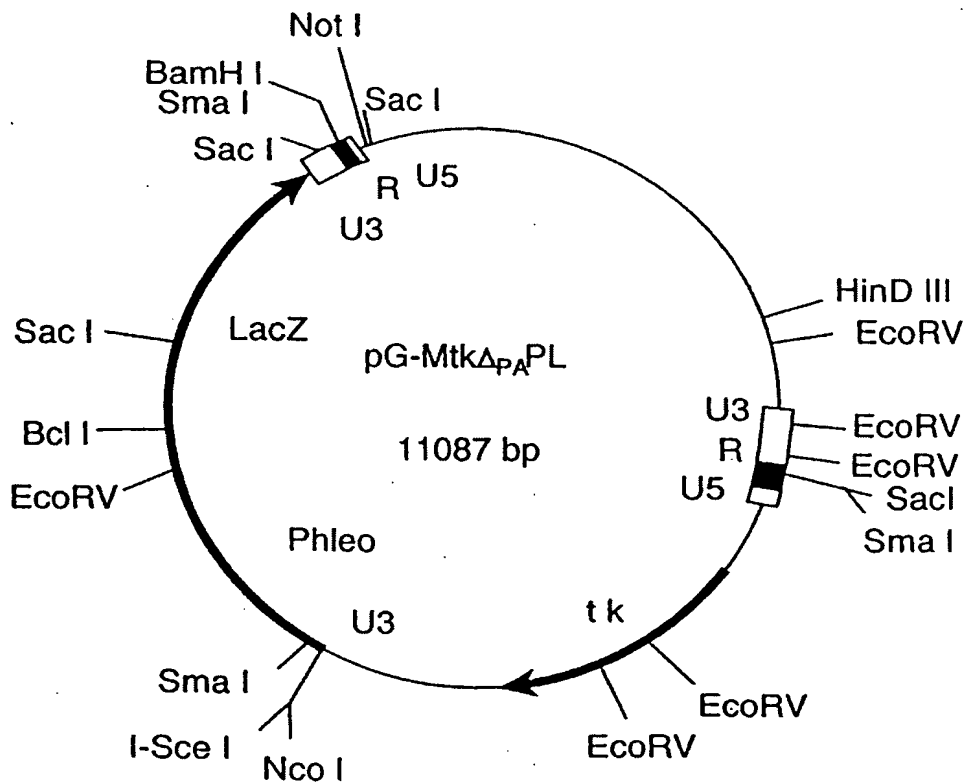


FIG. 33